Knowledge Management approaches to model pathophysiological mechanisms and discover drug targets in Multiple Sclerosis

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Abdul Mateen

aus

Sukkur, Pakistan

Bonn 2018

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

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

2. Gutachter: Prof. Dr. Joachim L. Schultze

Tag der Promotion: 04.05.2018

Erscheinungsjahr: 2019

رَّبِّ زِدْنِى عِلْمًا

My Lord! Increase me in knowledge.

Abstract

Multiple Sclerosis (MS) is one of the most prevalent neurodegenerative diseases for which a cure is not yet available. MS is a complex disease for numerous reasons; its etiology is unknown, the diagnosis is not exclusive, the disease course is unpredictable and therapeutic response varies from patient to patient. There are four established subtypes of MS, which are segregated based on different characteristics. Many environmental and genetic factors are considered to play a role in MS etiology, including viral infection, vitamin D deficiency, epigenetical changes and some genes.

Despite the large body of diverse scientific knowledge, from laboratory findings to clinical trials, no integrated model which portrays the underlying mechanisms of the disease state of MS is available. Contemporary therapies only provide reduction in the severity of the disease, and there is an unmet need of efficient drugs. The present thesis provides a knowledge-based rationale to model MS disease mechanisms and identify potential drug candidates by using systems biology approaches. Systems biology is an emerging field which utilizes the computational methods to integrate datasets of various granularities and simulate the disease outcome. It provides a framework to model molecular dynamics with their precise interaction and contextual details.

The proposed approaches were used to extract knowledge from literature by state of the art text mining technologies, integrate it with proprietary data using semantic platforms, and build different models (molecular interactions map, agent based models to simulate disease outcome, and MS disease progression model with respect to time). For better information representation, disease ontology was also developed and a methodology of automatic enrichment was derived. The models provide an insight into the disease, and several pathways were explored by combining the therapeutics and the disease-specific prescriptions. The approaches and models developed in this work resulted in the identification of novel drug candidates that are backed up by existing experimental and clinical knowledge.

Acknowledgements:

First and foremost, I would like to thank my research advisor, Prof. Dr. Martin Hofmann-Apitius, of Bonn University, for providing the intellectual support during my research. His guidance and valuable feedback helped me to become a mature scientist. I extend my gratitude to Dr. Luca Toldo, who facilitated me to conduct research in the department of Knowledge Management at Merck KGaA. His continuous support and enthusiastic encouragement broadened my horizons with industrial research. His mentoring transformed me into a curious person and refined my personality, both personally and professionally. I would like to thank Dr. Natalia Novac for being there in difficult times and solving problems with her excellent managerial skills. Furthermore, I would like to thank Prof. Joachim L. Schultze for his willingness to be the co-supervisor of my thesis. I am also thankful to Prof. Dr. Thorsten Lang and Prof. Dr. Diana Imhof for being part of the Ph.D. evaluation committee.

I would like to acknowledge the hard work of scientific collaborators from the Drug Discovery Department of Catania University, particularly Prof. Dr. Santo Motta and Dr. Francesco Pappalardo. Their support and excellent collaboration resulted in many publications. I would also like to acknowledge the hard work of Corri Johnson, who helped me with proofreading and improving the quality of my thesis. I would specially like to thank my parents and family for being there during stressful times. Their inspirational words always guided me through tough times.

Publication list

1. Rajput AM, s TM, Haase P, Scheer A, Toldo L: **SBML2SMW Links Systems Biology, Text Mining and Semantic Web**. Edited by Bichindaritz I, Perner P, s GR, Schmidt R. IBaI Publishing; 2012:177–184.

2. Rajput AM, Gurulingappa H: **Semi-automatic Approach for Ontology Enrichment Using UMLS**. Procedia Comput Sci 2013, **23**:78–83.

3. Rajput A-M, Pennisi M, Motta S, Pappalardo F: **OntoFast: Construct Ontology Rapidly**. In Knowledge Engineering and the Semantic Web. Edited by Klinov P, Mouromtsev D. Springer International Publishing; 2014:237–241. [Communications in Computer and Information Science, vol. 468]

4. Rajput A: **If It's on Web It's Yours!** In Trends in Practical Applications of Agents and Multiagent Systems. Volume 221. Edited by Pérez JB, Rodríguez JMC, Fähndrich J, Mathieu P, Campbell A, Suarez-Figueroa MC, Ortega A, Adam E, Navarro E, Hermoso R, Moreno MN. Springer International Publishing; 2013:189–192.

5. Pennisi M, Rajput A-M, Toldo L, Pappalardo F: **Agent based modeling of Treg-Teff cross regulation in relapsing-remitting multiple sclerosis**. BMC Bioinformatics 2013, **14** (Suppl 16):S9.

6. Gurulingappa H, Toldo L, Rajput AM, Kors JA, Taweel A, Tayrouz Y: **Automatic detection of adverse events to predict drug label changes using text and data mining techniques**. Pharmacoepidemiol Drug Saf 2013, **22**:1189–1194.

7. Fuller JC, Khoueiry P, Dinkel H, Forslund K, Stamatakis A, Barry J, Budd A, Soldatos TG, Linssen K, Rajput AM, HUB Participants: **Biggest challenges in bioinformatics**. EMBO Rep 2013, **14**:302–304.

8. Gurulingappa H, Rajput AM, Roberts A, Fluck J, Hofmann-Apitius M, Toldo L: **Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports**. J Biomed Inform 2012, **45**:885–892.

9. Gurulingappa H, Mateen-Rajput A, Toldo L: **Extraction of potential adverse drug events from medical case reports**. J Biomed Semant 2012, **3**:15.

10. Malhotra A, Gündel M, Rajput AM, Mevissen H-T, Saiz A, Pastor X, Lozano-Rubi R, Martinez-Lapsicina EH, Zubizarreta I, Mueller B, Kotelnikova E, Toldo L, Hofmann-Apitius M, Villoslada P: **Knowledge Retrieval from PubMed Abstracts and Electronic Medical Records with the Multiple Sclerosis Ontology**. PLoS ONE 2015, **10**:e0116718.

11. Pappalardo F, Pennisi M, Rajput A-M, Chiacchio F, Motta S: **Relapsing-remitting Multiple Scleroris and the Role of Vitamin D: An Agent Based Model**. In Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics. New York, NY, USA: ACM; 2014:744–748. [BCB '14]

12. Free System Biology Collaborative Workbench: Design and implementation (NETTAB 2010)http://www.nettab.org/2010/progr.html

13. A tool for fast development of new ontologies (INCOB 2014)http://incob2014.org/assets/INCOB2014/Poster-Session-v2.pdf

Table of Contents

Glossary

CellDesigner: CellDesigner [1] is a state-of-the-art structured diagram editor for drawing gene-regulatory and biochemical networks. Its intuitive user interface helps draw diagrams in rich graphical representation with personalized design. Networks are constructed based on a state transition diagram proposed by Kitano et al. [2].

DrugBank: The DrugBank database [3] is a comprehensive online drug database. It contains detailed information about drugs and drug targets. Due to its extensive descriptions, it is considered as drug encyclopedia. DrugBank contains almost 10,000 drug entries and each one of them contains more than 200 data field. It is widely used by scientists, students and the general public.

Ingenuity Pathway Analysis: Ingenuity Pathway Analysis (IPA) [4] is a life science knowledge base with powerful analytical and search tools. It integrates, stores and analyzes various types of biological data and help users identify new targets or candidate biomarkers. IPA has broadly been adopted by the life science research community and is cited in thousands of articles.

KNIME: KNIME [5] is an open source, easy to use graphical user interface workbench for different data analytics processes. It provides a broad range of nodes and plug-ins to connect to data preprocessing, connecting to web services, run scripts and execute external applications within the workbench.

Luxid: Luxid [6] is a commercial text mining system. It allows to mine text from various sources and with different skill cartridges (dictionaries) or patterns. The system also has ontology manager and semantic content enrichment features. The workflow option allows automation of certain tasks thus making it one of the favorite tools of text mining community.

MediaWiki: MediaWiki [7] is a open source, free software written in PHP. It was developed for Wikipedia but now available for any other use. There are several wiki websites running based on MediaWiki. The organization which owns MediaWiki is WikiMedia Foundation. MediaWiki has large variety extensions which extend its functionality. It is interoperable, robust and very stable.

MIRIAM: Minimum Information Required in the Annotation of Models (MIRIAM) [8] is a standardized set of metadata developed by the SBML community to facilitate the unified curation process of biological systems. The set of guidelines can be used with any structured format, allowing different groups to collaborate and share resulting models.

MySQL: MySQL [9] is an open source relational database management system (RDBMS) based on Structured Query Language (SQL). It is owned by Oracle Corporation.

Netlogo Language: NetLogo [10] is a programming language and integrated modeling suite totally oriented and devoted to agent-based modeling. It is free, open source, and developed in Java by Uri Wilensky in 1999 and it has been continuously updated ever since. It features an extensive documentation, multiple tutorials and a worldwide community that furnishes great support. NetLogo represents a good choice to simulate multi-agents, networks and complex dynamical systems. Many scientific articles have been published using NetLogo.

OntoFast: OntoFast [11] is an application that speeds up the development of new ontologies. It provides an easy to use and convenient interface that facilitates to build an ontology with associated metadata in short time. The output of the program can be easily opened and then used into a standard ontology editor like Protégé.

Protégé: Protégé [12] is a free, open-source ontology editor. It was developed by the Stanford Center for Biomedical Informatics Research. Protégé provides a suite of tools to develop ontologies, domain models, and knowledge-based applications. It has many plug-ins which extend its functionality.

SBML2SMW: SBML2SMW [13] is CellDesigner plug-in that stores CellDesigner models in a Semantic MediaWiki format. This plug-in allows extracting CellDesigner model information, storing this information to a Semantic Mediawiki server and context-sensitive restoring and integration of this information in a CellDesigner model. The application consists of two parts: The CellDesigner plug-in itself which directly communicates with the CellDesigner and so has access to the CellDesigner models and the Translation Server which receives the extracted information and translates it into Semantic Mediawiki syntax and stores it there.

SCAIView: SCAIView [14] is a semantic search engine for life sciences. It processes large volume of text to facilitate the quick identification life science concepts. The backend system works on technologies such as text mining and semantic web to provide a refine search results.

UMLS: Unified Medical Language Systems (UMLS) [15] is a very large repository of medical concepts which integrates and streamlines many health vocabularies to enable interoperability among them. UMLS has more than 100 source vocabularies and it has been reported that the 2009AB release of the UMLS Metathesaurus contained 2,120,271 biomedical concepts and 5,305,932 unique terms [16]. Two areas of its usage are in electronic health records software development and by health-related language translators. UMLS deals with the complexity of different biomedical concepts by assigning a unique identifier to them, called a Concept Unique Identifier (CUI).

1. Chapter 1: Introduction to Scientific Challenges

1.1 Complexity of the human brain:

The human brain is who we are; it is our core, it makes us, it controls all of our physical systems, and consciousness. It can be considered as a very advanced-level programming language, as by definition a programming language is a formal set of words designed to communicate instructions to a system [17]. The brain communicates instructions to different systems of the body and reprograms instructions in real time, based on the feedback. It is a processor which takes input, processes it, and delivers output. It stores our short and long term memories. It receives feedback, sends orders, and decides the actions based on the collected information. Multiple sensors bring information to the brain so it can make decisions about anything which happens around it. The brain remains in darkness but still makes us feel brightness. It is in a silent place but allows us to hear loudness. It allows us to taste, smell, and be happy during certain events, or performing certain actions. Furthermore, the brain does not need eyes to let you see; for example, dreams can be experienced and visualized whilst the eyes are closed. In fact, we live our life in the brain; clinical death is only announced when the brain ceases to function. In addition, the brain is one of the few organs which remains functional all the time and does not rest like any other body part. It is only our brain which distinguishes us from other animals; as we extend the possibilities and solve the mysteries of universe.

The human brain is considered to be the most complicated structure known in the universe and the most complex organ of the body. It has been shown that the human brain has approximately 86 billion neurons and almost the same number of non-neuronal glial cells [18], although only 302 neurons are required for a living organism with a functional nervous system (C. elegans have only 302) [19]. Each neuron can be connected to as many as 10,000 others, thus connections could potentially reach up to 100 trillion. The existence and functionality of the plethora of molecules in the human brain are still ambiguous and largely unknown [20]. The enormous numbers of cells and connections make it further difficult organ to study.

Since the human brain directs all the other parts of the body to react to different environmental stimuli, the interplay between environment and genetic information is crucial for brain activity. Therefore, the role of environment is crucial to its well-being. The interaction of both of these factors is called gene-environment interaction. One interesting and simple example of the role of the environment in our health is the risk of skin cancer in fair skinned people being significantly higher than that of dark skinned people, through prolonged exposure to sunlight [21]. In the following chapters, the crucial role of environmental factors (specifically vitamin D deficiency) in some aspects of brain diseases as well as in brain deterioration will be discussed. Often the environmental role is in chronological order as most of victims are elderly people.

Neurodegenerative diseases are characterized by the degeneration of cells in the nervous system. It is a broad term used to define different diseases which have similar characteristics; they act on neurons to degrade or destroy them. Almost all of the neurodegenerative diseases impair brain functionality and they are complicated to treat because of our limited knowledge of brain functions. Neurodegenerative diseases are an area of major concern to healthcare providers, since they are a burden on the social system due to the occurred expenses during treatment, and they make patients handicapped and dependent in a chronic fashion. Many initiatives have been introduced to understand brain functionality and to better cope with neurodegenerative diseases. Human Brain Project [22], BrainInitiative [23], Allen Brain Atlas [24], Blue Brain Project [25], and BrainMaps [26] are a few among countless projects.

The ongoing research and number of publications associated with neurodegenerative diseases are increasing at a higher rate (Figure 1.1) but the need for new therapies for different brain disorders remains largely unmet. Due to the brain’s complexity, there have been a number of failed attempts to create a drug which can combat the diseases impacting the brain. Many tests show promising results in model organisms but not in humans. Furthermore, the model organisms only represent few aspects of the human brain, thus limiting the scope of research and drug development.

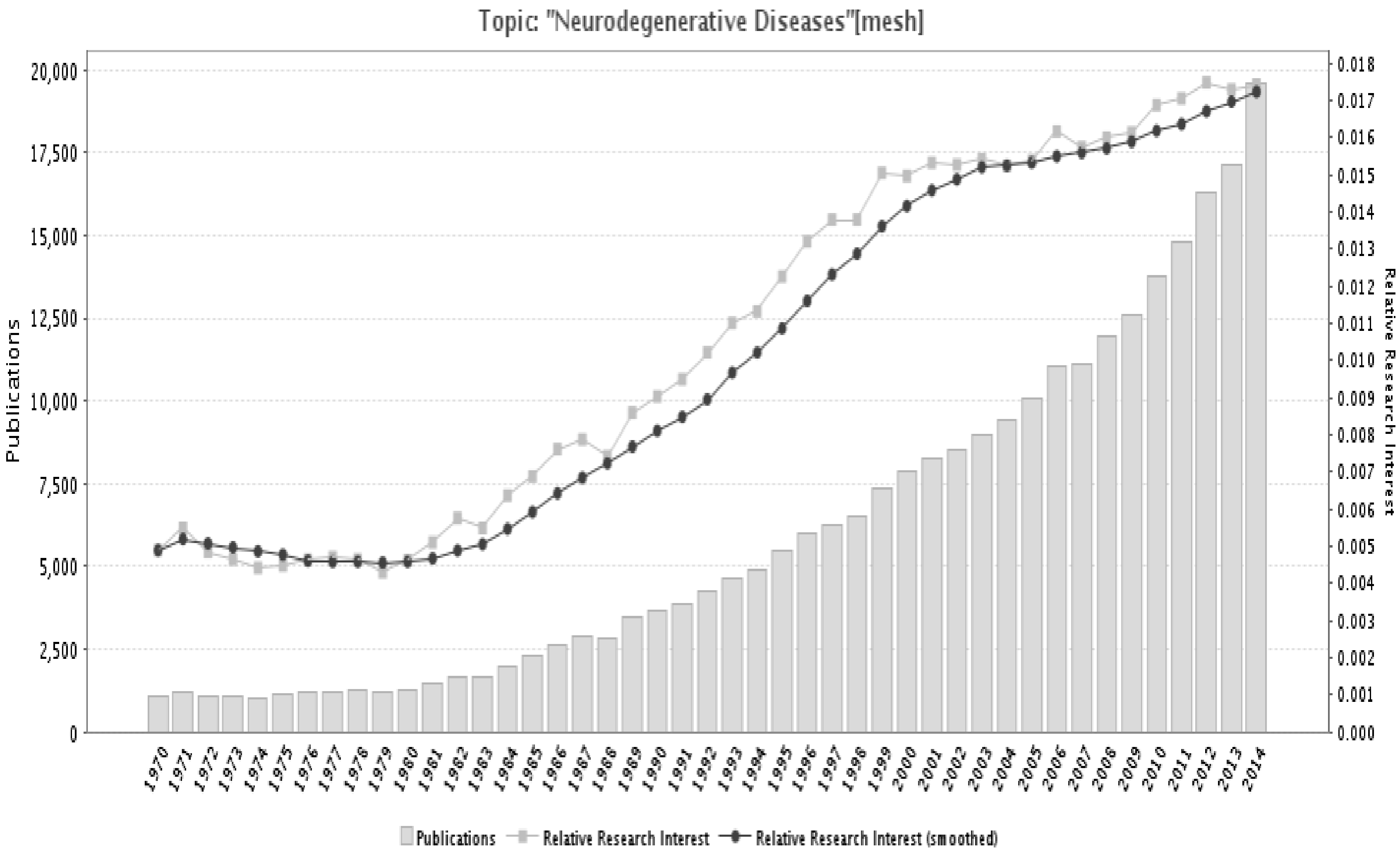


Figure 1.1: Publications record indexed in PubMed with the Mesh term "Neurodegenerative Diseases" over time.

The massive amount of scientific literature available for different brain disorders is not useful enough to find a cure of neurodegenerative diseases. The quest to transform “data into knowledge and wisdom” could be advanced by changing the ways we look into and interpret the literature. The following chapters will demonstrate the use of various knowledge management approaches, in order to tackle the daunting task of finding hidden pearls of knowledge associated with a specific disease, namely Multiple Sclerosis. Although these methods will be applied here to Multiple Sclerosis, they can be applied to any other neurodegenerative disease.

1.2 Research motivation:

After the successful eradication of common infectious diseases, the new battleground for researchers is the study of diseases whose causative agents are either unknown or involve more than one factor. Brain diseases are some of the most challenging due to the complexity of the organ, its isolation from the environment, and the roles played by genetic and environmental factors in these diseases. Most of the diseases happen to occur in a chronological order demonstrating the role of genetic and environmental factors’ interplay. In addition to brain being complex, the diseases are prone to be more difficult to study as the brain is not exposed directly to the environment. Therefore, special equipments are needed to investigate them.

Despite the modern technology and discovery of advanced molecular biology techniques, researchers are still struggling to find a cure for neurodegenerative diseases. One of the major neurological diseases which remains incurable is Multiple Sclerosis (MS). MS is one of the most common disorders of central nervous system, affecting between an estimated 2.3 to 2.5 million people worldwide [27].

MS starts at an early age, renders people handicapped through lifelong progression of the disease, and many complications are associated with it. The major consequences of having MS are neuronal damage and the unpredictable course as the disease progresses. It causes significant disability in patients and has a considerable influence on the personal life of the patient, with regards to social impact, cost, and quality of life. The cost of living with MS is significantly higher than other brain diseases. Figure 1.2 shows the comparative survey of costs associated with different disorders in European countries.

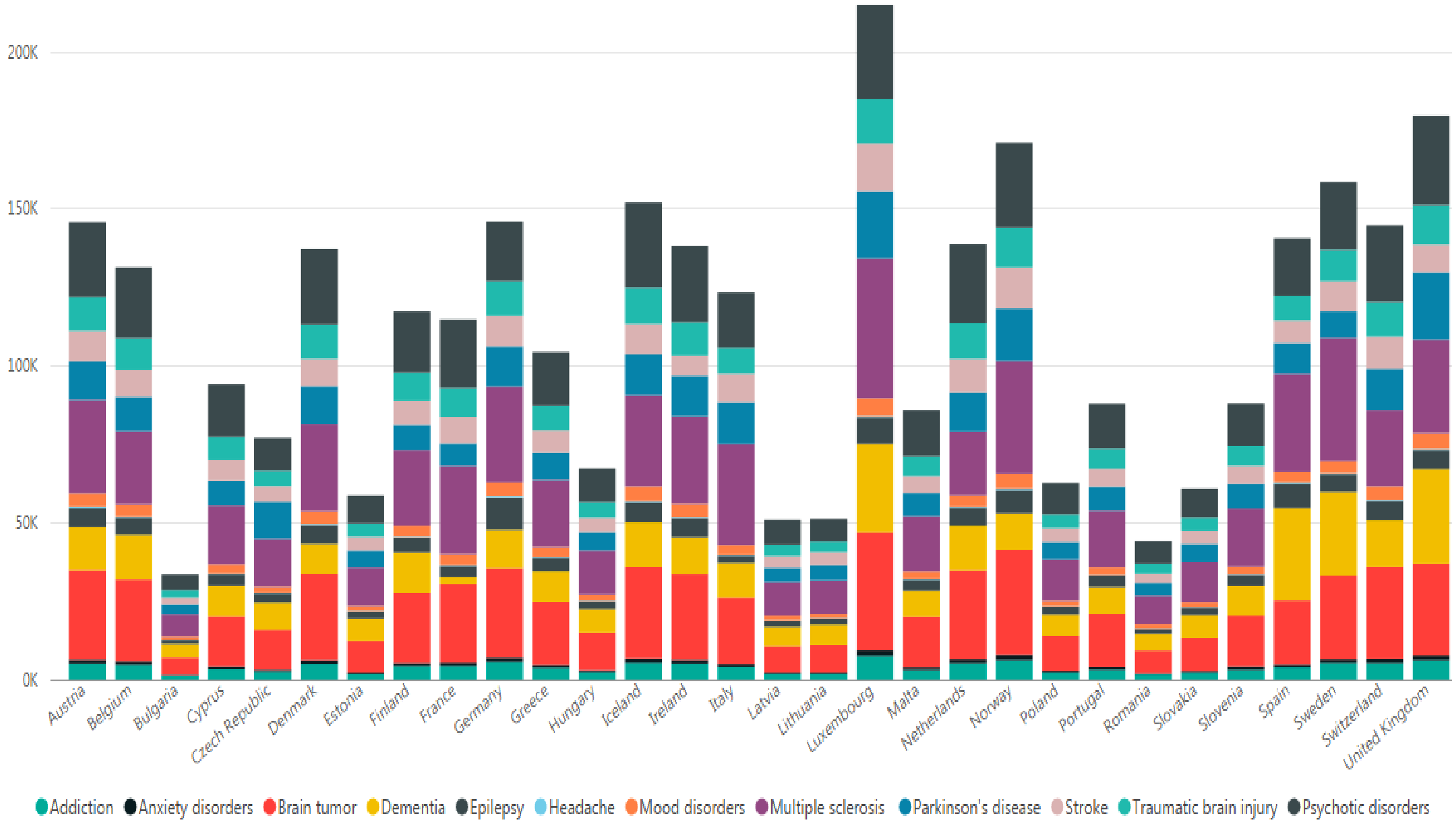


Figure 1.2: Cost per person with a disorder of the brain in different European countries. Purple colour shows the data of Multiple Sclerosis and it can be seen that in most of the European countries the cost of MS treatment is higher than any other diseases. edited version of [10].

The incurred costs of MS treatment are not only higher than any other disease, but indirect costs and other non-medical costs also make it one of the expensive diseases to treat (Figure 1.3).

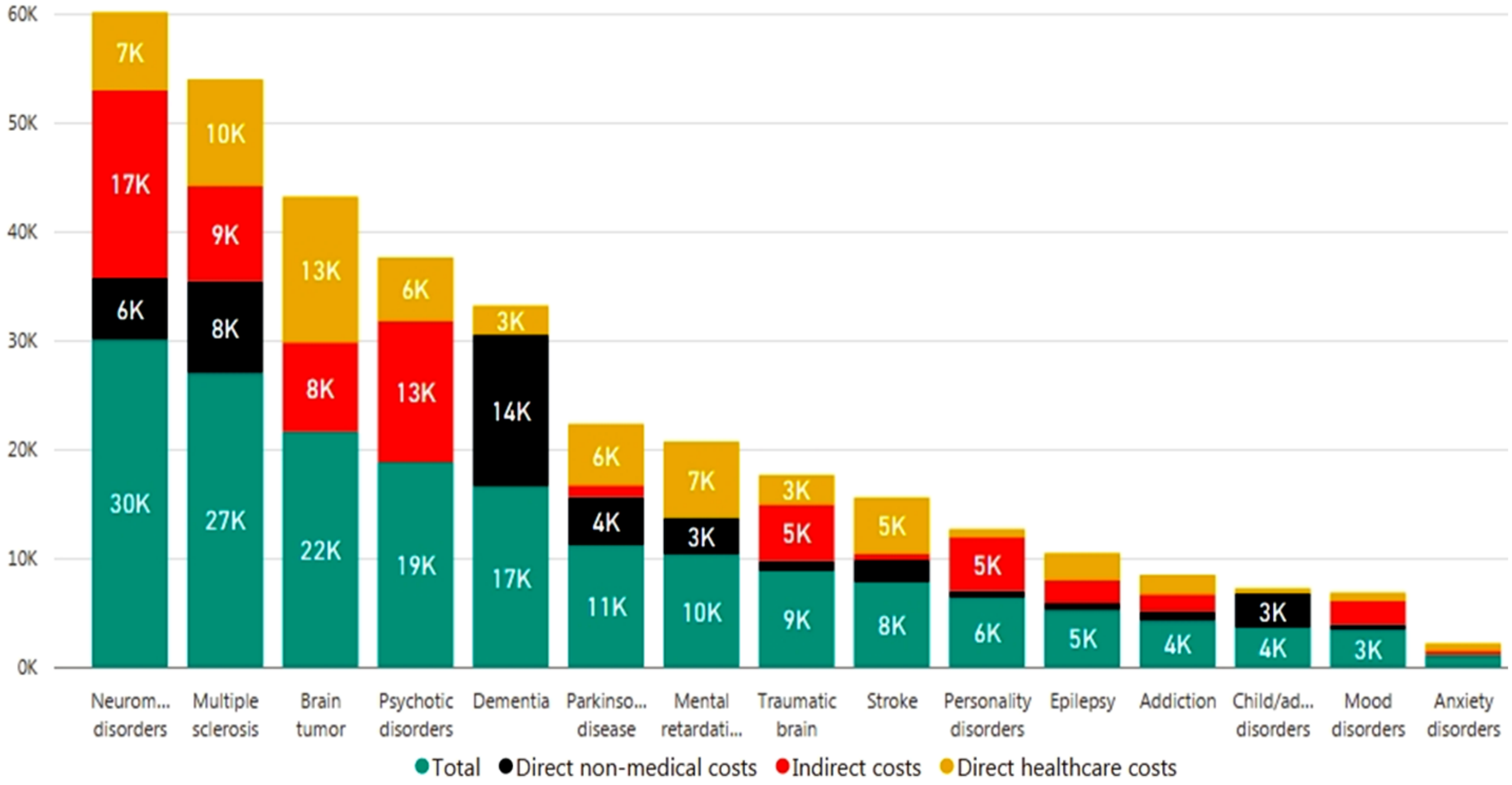


Figure 1.3: Non-medical/indirect cost per person (in €) of different brain disorders comparison, taken from [10].

1.3 Research goal:

The goal of the research work is to foster drug discovery by:

- Extracting relevant information from public and proprietary data (e.g. clinical trials), in order to unravel the hidden pieces of knowledge and intertwine them; this could reveal the underlying mechanism of different disease phenomena.
- Modeling knowledge based on molecular interaction maps and developing models of MS in time dependent manner with biomarkers involved in the different stages of the disease.
- Simulating different aspects of MS disease by using various modeling approaches which could reproduce similar patient outcomes after an in-silico experiment has been performed.
- Increasing knowledge reach by improving current tool sets. This includes gaining access to datasets which are available in languages other than English.

1.4 Thesis outline:

Following is a brief outline of the upcoming chapters and respective papers linked to them.

Chapter 2 discusses about the complexity of MS, risk factors, processes involved in the disease, disease subtypes, biomarkers associated with the disease, and available therapies currently in the market.

Chapter 3 focuses on various systems biological approaches to model MS, their limitations, and their role in drug discovery. Molecular interaction maps, systems biology languages, and agent based modeling are discussed in details. Systemic review of the current literature is also discussed.

Chapter 4 focuses on the methodology part of all the work done in this thesis. In this chapter, various approaches about information retrieval and modeling have been described. The methodology part was divided into two parts; first part discusses the foundational work for developing the MS models and all information retrieval work while second part focuses on the modeling work.

Chapter 5 discusses the results and findings of the thesis. The main outcomes of the thesis are:

- A molecular interaction model of the MS disease
- An automated methodology to enriching and/or translating any ontology

- A disease specific ontology which has been translated and used to mine Electronic health records
- Time series patterns of MS progression and speculation of combination therapy

Chapter 6 discusses the forthcoming possibilities of the work and limitations faced during the work. It also describes the outcome of the research and their application.

In the following section the outline of thesis' result section is given with all the publications under relevant topics. In addition, a paragraph of my contribution is also provided.

Information retrieval and representation:

- Ontology Development [28]:
  - This paper describes the work of processing large amount of information and information representation. In order to retrieve useful information from scientific literature and electronic medical records (EMR) an ontology specific for Multiple Sclerosis (MS) has been developed. The paper relates to the goals mentioned in the first chapter "Extracting relevant information from public and proprietary data" and "Increasing knowledge reach by improving current tool sets". The MS Ontology was created using scientific literature and expert review under the Protégé OWL environment. The MS Ontology was integrated with other ontologies and dictionaries (diseases/comorbidities, gene/protein, pathways, drug) into the text-mining tool SCAIView, a tool developed by Fraunhofer Institute for semantic search. MS ontology has also been used to analyze the EMRs from 624 patients with MS in order to identify drug usage and comorbidities in MS. The challenges faced were that no structured knowledge was available about MS, and limited available tools to mine datasets available in languages other than English. The challenges were dealt by developing a disease specific ontology and a methodology to translate any ontology.
  - My contribution for the work was; collecting all the concepts, developing the structure of ontology, mapping the concepts to UMLS and MeSH, enriching the ontology with the synonyms, definitions and identifiers, and translating the ontology into Spanish so it can be used for Spanish EMR dataset shared by collaborators.

- Automated data retrieval from UMLS[29]:
    - This paper describes a methodology which enriches ontologies in significantly less time and with an automated way. This paper relates to the thesis because ontology enrichment was the foundation of ontology development and ontology was developed to have a unified source of knowledge for modeling the disease. One of the challenges faced during the ontology development was enriching the terms with relevant concepts. Ontology enrichment is a process of embedding metadata associated with concepts described in the ontology. Manual information retrieval and enrichment process is labor-intensive and time-consuming as each concept is unique and has domain specific meanings. An approach to deal with this problem is to have a unified resource and an automated solution.
    - My contribution for the work was; everything, namely; building the workflow, installing and configuring UMLS locally, automating the procedure, running the queries, testing the output, cleaning the output, and writing the publication. The second author only reviewed the publication and tested the system.

- Automated Ontology Translation from UMLS:
    - Not published yet.
    - All the work done by me.

- OntoFast [11]:
    - This publication describes a tool named OntoFast, which provides an easy to use interface for ontology development. Ontology development is a time-consuming task and with the help of this tool, ontology can be developed in few days. The challenge faced was the complex structure of OWL and the way it stores the embedded knowledge. The task of the tool development was completed after thorough research and investing some time.
    - My contribution for the work was; proposing the idea, monitoring the development, providing all the relevant knowledge, testing, exclusively using the tool, and writing the manuscript.

- Corpus generation from web to analysis the content and applying NLP tools [30]:
    - This publication describes the approach of collecting clean dataset from public domains. This also relates to the goals mentioned in the first chapter “Extracting relevant information from public and proprietary data”. The challenge faced was that required data was not easily accessible and concealed under many layers. The challenge was dealt by providing a methodology which can process the data under several layers of web pages and it can be applied to any other website.
    - My contribution for the work; I am the only author.

- SBML2SMW, Transforming Systems Biology knowledge into MediaWiki Pages [13]:
    - This publication[13] describes about the connected framework established for integrating, storing and reusing the scattered knowledge of different domains e.g. Systems biology tools, Text mining systems, and semantic web applications. This relates to one of the goals mentioned in the first chapter, “Extracting relevant information from public and proprietary data”. The challenges faced were that different domains of knowledge provide different data format and often the outputs are not interoperable. We linked the two very popular technologies, MediaWiki and Systems Biology Markup Language by developing a free plug-in which bridges the gap between them and increases knowledge reach.
    - My contribution for the work was; proposing the idea, monitoring the development, testing, exclusively using the system, and writing the manuscript.

Modeling of MS disease:

- Molecular interaction map of MS [31]:
    - Molecular interaction maps (MIMs) are interaction maps of molecules that are involved in a biological function. The map of MS is published at Payao website [31]. MS map is one of its kind as it contains text extracted from scientific literature under each edge with PMIDs. The map was developed by manually reading research papers. Challenges were that manual reading is time-consuming and map visualization is an issue after map has developed to a certain extent. Challenges were dealt by using various semantic software and running queries to retrieve the required knowledge.

    - My contribution for the work was; I am the only author.

- Time series of MS disease progression and combinatorial therapy:
    - Not published yet.
    - All the work done by me.

- Agent based model of Treg-Teff interplay and its role in RRMS: The Model and Simulation [32]:
    - This paper describes the interplay between Treg-Teff and the role in relapsing remitting subtype of MS. The work was done by using Agent based modeling technique and NetLogo application was used. The model shows results obtained from eight randomly chosen individuals. They were genetically predisposed mimicking absence and presence of malfunctions of Teff-Treg cross-balancing mechanism. The presented model allows to capture the essential dynamics of relapsing-remitting MS despite its simplicity. It gave useful insights that support the hypothesis of a breakdown of Teff-Treg cross balancing mechanisms.
    - My contribution for the work was; shared the knowledge about MS and review the publication.

2 Chapter 2: Complex Biological Mechanisms Underlying Multiple Sclerosis

2.1 Multiple sclerosis:

Multiple Sclerosis (MS) is one of the most common chronic inflammatory diseases which affects the central nervous system and causes its deterioration. MS is also known as disseminated sclerosis or encephalomyelitis disseminata. It usually begins between age 20 to 50, and the ratio between males and females is 1:2 [33]. It causes sclerosis, which is a Greek word meaning “hardening” and the term refers to myelin sheath hardening or the formation of lesions on the myelin. In this disease, lesions mostly occur in the white matter of the brain and spinal cord. MS is a neurodegenerative disease which causes hindrance in axon communication by injuring the myelin sheath, which results in different signs and symptoms, including physiological and mental problems [34]. Besides the patient’s personal life dilemma, MS also causes a significant social burden on the healthcare system. It has been estimated that between 2 to 2.5 million people are affected by MS, and the cost associated with it is significantly higher as compared to other neurological diseases. This higher cost is mostly due to the chronic nature of the disease. It is more prevalent in Europe, Canada, and the US as compared to rest of the world. Table 2.1 shows the natural history of MS, percentage of women affected, and onset age.

	Italy	**Canada**	**France**	**Canada**	**Sweden**	**USA**
Number of patients	1463	2837	1844	1099	255	201
% Women	67.2%	70.4%	64.4%	65.7%	60.0%	69.7%
Onset age (all)	28.3 years	30.6 years	31.0 years	30.5 years	–	31.2 years
(standard deviation)	(9.0)	(10.0)	(9.7)	(9.9)	–	–
< 15 years	3.8%	–	–	–	–	–
< 20 years	18.7%	9.6%	11.7%	12.0%	–	–
> 40 years	11.6%	13.8%	20.6%	20.1%	–	–
> 50 years	1.9%	4.7%	5.9%	–	–	–
Relapsing onset (%)	86.1%	87.6%	84.7%	81.2%	85.9%	94.5%
Earliest onset	–	–	5 years	–	–	12 years
Latest onset	–	–	62 years	–	–	58 years
Mean age relapsing onset	–	–	29.5 years			
Mean age progression onset						
SPMS	–	49.0 years	39.5 years	40.1 years	–	–
PPMS	–	41.0 years	39.3 years	38.6 years	–	–
Time to EDSS = 4, median	–	–	8 years	–	–	–
Relapsing onset	–	–	11 years	–	–	–
Progressive onset	–	–	0 years	–	–	–
Time to EDSS=6, median	–	28 years	14 years	15 years	18 years	26 years
Relapsing onset	–	–	23 years	–	23 years	28 years
Progressive onset	–	–	7 years	–	6 years	7 years

Table 2.1: Summary of natural history of Multiple Sclerosis in different countries, edited version of [35].

MS is a complex disease for numerous reasons. Its etiology is unknown, the diagnosis is not exclusive, the disease course is unpredictable, and therapeutic response varies from patient to patient. The etiology of disease is still ambiguous despite the fact that the disease was firstly described in 1868 [36], and medical science has advanced significantly since then. There are many hypotheses about the brain regions mainly involved [37] and disease causation. Viral infections [38], epigenetical changes [39], mitochondrial defects [40], Vitamin D deficiency [41], and the role of genetics [42] have been proposed to play a role in disease etiology. The diagnosis is also very challenging; physicians must rule out other ailments by asking different questions and taking tests, as many other medical conditions share similar symptoms. Even the expert physicians can diagnose the disease correctly only 90-95 percent of the time [43]. There is no single test which can be conclusive about the diagnosis. The disease course of MS can be acute or chronic. As the disease is highly unpredictable, the victims normally do not know which organ or system will be affected next. The disease course variation between individuals means that treatments may not affect all patients in a similar manner. This includes one of the most effective drugs against MS, interferon beta.

What causes MS is still an open question, however there are factors which are considered relevant and are associated with the frequency of disease occurrence. The deficiency of Vitamin D, infection agents (specially Epstein-Barr Virus - EBV) and presence of some genetic variations are among them and are associated with a higher probability of disease occurrence [44]. It is also very interesting to know that MS is more common in people who live farther from the equator (Figure 2.1) than in those who live near it, though there are some exceptions [44][45]. Even though some genes show significant association with MS and females display a twofold increased likelihood to develop the disease, it is not considered to be a hereditary disease. The risk of having MS increases depending upon the closeness of the relation of the diseased person [46]. For instance, identical twins have a 30% probability of developing the disease if one of them is affected. In comparison, non-identical twins have only a 5% probability, and siblings have a probability of 2.5%. In addition, if both parents are affected, then the risk of having children with MS increases by 10 times compared to general population [33].

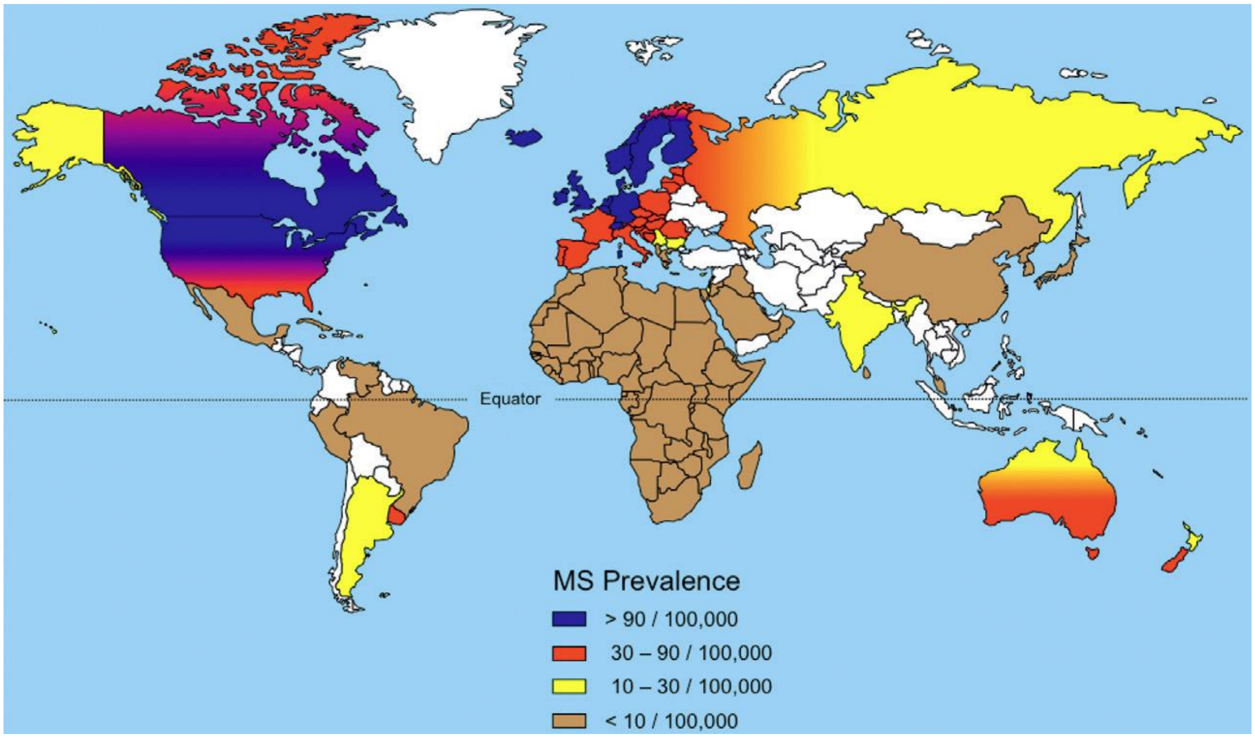


Figure 2.1: Global distribution of MS prevalence. MS is more common in people who live farther from the equator, taken from [35].

2.1.1 Relapse and remission:

Relapse is a term which defines a period of worsening of disease activity; it includes the development of new symptoms or reoccurrence of previous symptoms, with or without increased severity. Remission is defined as complete or partial recovery of the symptoms, following relapse. Some factors can halt or trigger a relapse, e.g. pregnancy. Relapse is less likely during the gravid vs. non gravid state, and is especially unlikely during the third trimester [47]. Factors which could increase the risk of having a relapse are infections and other ailments which include fever, cough, rhinorrhea, nausea, and diarrhea as symptoms [48–51].

2.2 Risk factors of multiple sclerosis:

There are many risk factors associated with MS; some are proven to have a role in disease pathogenesis. Others are hypothesized but do not have sufficient data available to prove their significant influence on the disease etiology. Here we will discuss some of the factors as shown in the figure 2.2.

The genes most commonly associated with MS are the HLA genes, a known fact for more than 30 years [52]. Interestingly, it has been found recently that HLA genes have a role in Vitamin D gene expression [53]. Furthermore, the same group of genes are also associated with other autoimmune diseases such as diabetes mellitus type 1 and systemic lupus erythematosus [52]. DR15 and DR6 have been found to be more consistent alleles in MS disease association [44]. On the contrary, some other alleles such as HLA-c554 and HLA-DRB1*11 have been shown to have a protective role [44]. In addition, HLA-DR15 occurrence within the Caucasian race also increases the likelihood of the disease. There are also certain haplotypes which are linked to an increased likelihood of MS disease, e.g. IL7RA and IL2RA. Primary oligodendropathy and some of the single nucleotide polymorphisms (SNPs) are also linked with disease etiology [54,55].

Besides genetic predisposition, viral infections such as EBV also play an important role in MS prevalence. It has been documented that an early age infection of EBV has a protective role against the disease and late infection is considered as a risk factor. Individuals who have never been infected by EBV are at a reduced risk of getting MS as compared to those who got infected at younger age [44][56]. High Epstein–Barr nuclear antigen 1 (EBNA-1) titres are also considered as a risk factor in disease occurrence. The environment has also been considered to play a crucial role in MS disease etiology. Low exposure to sunlight (resulting low Vitamin D levels) and babies born in May are more prone to develop the disease as compared to babies born in November. It could be argued that it is also due to Vitamin D deficiency, as mother was less exposed to sunlight during pregnancy. Another important environmental factor is smoking, which is linked to an increase in the likelihood of disease.

The following single molecules are considered to play a significant role in disease etiology, and there are certain drugs which target these molecules and suppress the disease. For example, VLA-4 is one of the molecules which is suggested to play a role in disease. It has been proven by the fact that the drug Natalizumab reduces disease severity by targeting it. Other molecules which are considered to have a role in the disease are IFN-y, IL-17A, IL-23, Osteopontin, Complement system and ROS.

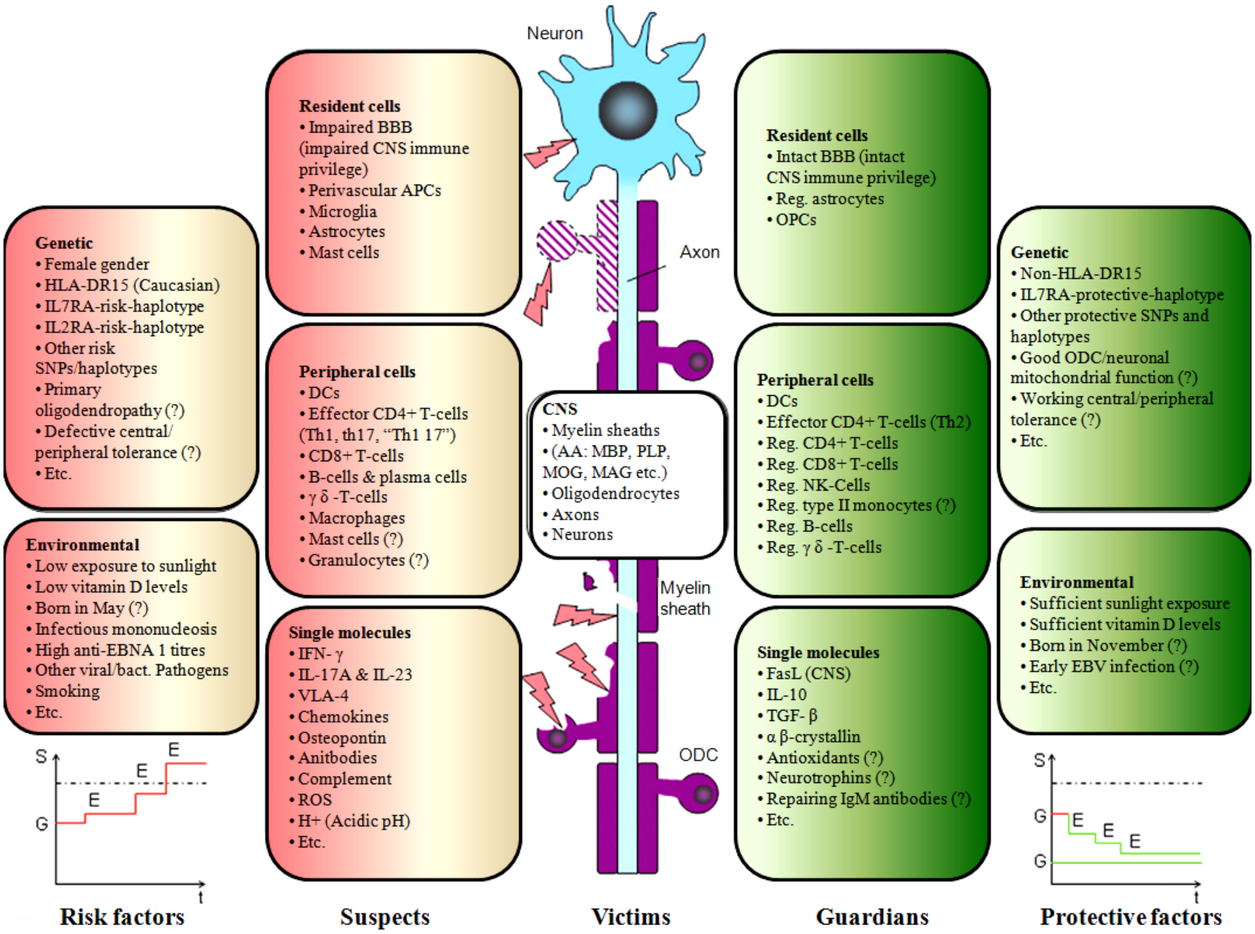


Figure 2.2: An overview of different factors involved in MS. Risk factors on the left (pink) increase susceptibility (S) to disease likelihood (dashed line). Strong genetic components (G) and environmental factors are part of them. Suspects are entities which are thought to be overall deleterious. In the centre of the image, the damage in MS has been shown; red pointers show the points of attack i.e. myelin sheath, oligodendrocytes, axons and neurons. Guardians are entities which are thought to repair the damage. Protective factors (green) decrease susceptibility to the disease, edited version of [35].

2.3 The three pillars of MS disease: Demyelination, neurodegeneration and autoimmunity:

2.3.1 Demyelination or myelin damage:

Myelin damage (also known as demyelination) is a term used to define a condition when the myelin sheaths of the axons become damaged, and the electrical pulse is lost during transportation. Different patterns of myelin damage have been documented, e.g. myelin stripping, dissolution of myelin sheath by invasion of macrophages [57] and binding of myelin fragments to vesicles of macrophages [58]. Demyelination causes signal deterioration and as a consequence, organs which are supposed to act on nerve signals stop responding. This can cause body systems to malfunction. This state has been considered as the disease onset, and is associated with relapsing remitting MS (RRMS).

2.3.2 Neurodegeneration:

Neurodegeneration is a broad term which describes the processes involved in the deterioration of the nervous system; it literally means nerves (neuro-) and destruction (-degeneration). Neurodegeneration in MS occurs when immune cells enter the CNS and cause neurotoxic effects. Lesion formation, which is a hallmark of MS, occurs when immune cells react with myelin protein [59,60] causing damage to nerve cells. There is a large body of literature available which supports the notion that neurodegeneration is a very important process in MS. Histological examination and postmortems of brains show the axonal damage and subsequent loss [61,62]. Progressive brain atrophy and reduction of N-Acetyl aspartic acid have also been documented extensively in MS [63,64]. In addition to neurodegeneration, it has been shown that immune cells also play a vital role in neuroprotection after entering the CNS [65,66]. It has been shown that there are two different mechanisms involved in neurodegeneration with respect to acute and chronic subtypes of MS. Acute MS is characterized by lesions infiltrated by macrophages and T-cells. In chronic MS, activated microglia cause axonal damage. The balance among different populations of T-helper cells could yield different outcomes in the disease; this concept is known as T-helper cell polarization. The concept as discussed in [67] suggests that TH1 and TH-17 are responsible for CNS inflammation and neurodegeneration, while a sub-population of T-helper cells known as TH2 cells seem to play a neuroprotective role. This hypothesis is supported by findings that Glatiramer Acetate and Statins promote TH2 polarization [68,69] to reduce the severity of the disease. In addition to TH cell populations, macrophages also have a similar pattern which results in a different cellular phenotype population [70]. M1 macrophages are considered to have a pro-inflammatory role, contributing to tissue destruction, while M2 macrophages have an anti-inflammatory role, contributing to tissue repair. Interestingly, Glatiramer Acetate can induce a population shift towards M2 phenotype in MS [71]. This implies that shifting cell phenotypes could be a possible treatment option for neurodegeneration.

2.3.3 Autoimmunity:

MS is a disease of the immune system in which autoimmunity is believed to play a crucial role. One of the strongest supportive arguments of this notion is that in the animal disease model of MS, experimental autoimmune encephalomyelitis (EAE) has immunopathological processes in the CNS which exhibit many aspects of MS [72]. Further supportive arguments include the presence of auto-antibodies [73], auto-reactive T-cells [74], anti-MOG antibodies [75], and MHC class II complexes with MBP peptides on antigen- presenting cells in MS lesions [76]. In addition, nearly all of the therapeutic agents given to suppress the MS disease activity act to effectively suppress the immune system, or otherwise alter its function [77]. Many of the genes involved in MS are often associated with other autoimmune diseases, supporting the theory that it is an autoimmune disease (Figure 2.3) [78]. These findings suggest that the immune system plays an important role in disease pathogenesis; however, whether these immunological processes are the primary cause of the disease or the reactions of some other stimulus is yet to be understood.

The primary players of the disease pathogenesis are the auto-reactive T-helper cells that attack the myelin basic protein (MBP) and cause damage to it. T-cells cross the blood-brain barrier via the adhesion molecules e.g. VLA-4. They become stimulated by antigen-presenting cells, release pro-inflammatory cytokines and cause demyelination along with B-Cells and macrophages [79–82]. Viral infections are suspected to play a key role in the activation of the T-cells' auto-reactivity by a process called molecular mimicry. Molecular mimicry is defined as when an immune cell cannot differentiate between bodily molecules and foreign particles due to significant structural similarities in the peptide sequences. Many viruses are homologous MBP including adenovirus, herpes simplex virus (HSV) and Epstein-Barr virus (EBV) [83]. Among these EBV has been shown prominently to be associated with immune system modulation and is considered to be an important factor in MS pathogenesis [84]. Due to multiple mimicry possibilities and different pathways' involvements, it has been challenging to identify a single causative agent of MS. It is also hypothesized that variations in MS disease phenotypes may be due to the different agents involved in the disease subtypes.

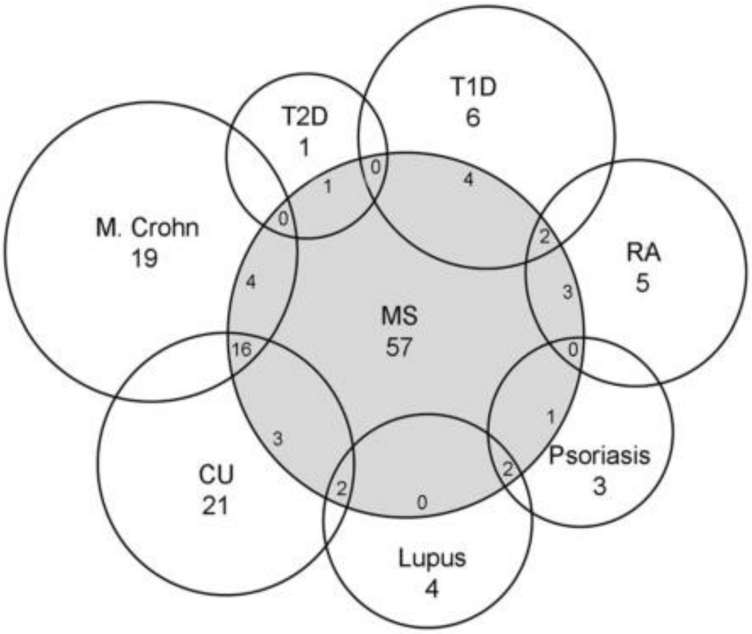


Figure 2.3: Venn diagram of genetic overlap among different autoimmune diseases: The external circle shows the number of shared genes between a disease and MS. The Internal circle shows the count of susceptible shared genes. Clockwise from upper right legends: T1D= Type 1 Diabetes, RA= Rheumatoid Arthritis, CU= Colitis Ulcerosa, T2D= Type 2 Diabetes, MS= Multiple Sclerosis. Taken from [85].

2.4 Heterogeneity in MS:

There are four major and established subtypes of MS, which are segregated based on different characteristics. MS commonly begins with RRMS and then progresses in severity. Clinically Isolated Syndrome (CIS) is also considered as a stage of the disease, but it will not be discussed here as some people who have experienced CIS may not develop MS [86]. Table 2.2 shows the characteristics of the different subtypes of the disease.

Type	Characteristics	Frequency at Disease Onset
RRMS	Clearly defined relapses followed by full or nearly full recovery of function. The disease does not progress in between attacks.	~85%
SPMS	Initial course of relapsing–remitting that is followed by progression, with or without occasional relapses or plateaus in function. The disease progresses over time.	~50% of people with RRMS will transition to SPMS within 15 years from diagnosis

PPMS	Gradual but continual worsening over time; some fluctuations but no distinct relapses	~10%
PRMS	Progressive course marked by distinct relapses, with or without recovery. The disease continues to progress between attacks.	5%

Table 2.2: MS disease subtypes and their characteristics, edited version of [94].

2.4.1 Relapsing remitting multiple sclerosis:

The standard description of Relapsing Remitting Multiple Sclerosis (RRMS) defined by clinicians [87] is as follows:

RRMS: "Clearly defined relapses with full recovery or with sequelae and residual deficit upon recovery; periods between disease relapses characterized by a lack of disease progression."

RRMS is the most common form of MS. It accounts for nearly 85% of all MS cases [88]. In this subtype of the disease, patients face recurring instances of relapses which fade away after some time. For most of the people this is the start of disease, as the diagnosis is usually made when the patient experiences something distressful like double vision, loss of balance or thinking problems. One of the hallmarks of this form of MS is that it does not progress between the attacks. Figure 2.4 shows the progress pattern in RRMS.

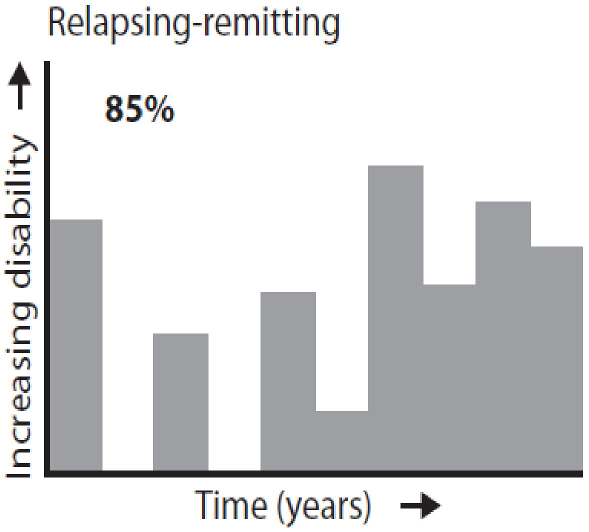


Figure 2.4: Disease progression pattern in RRMS, number of percentage shows the frequency of disease subtype.

2.4.2 Secondary progressive multiple sclerosis:

The standard description of Secondary Progressive Multiple Sclerosis (SPMS) defined by clinicians [87] is as follows:

SPMS: "Initial RR disease course followed by progression with or without occasional relapses, minor remissions, and plateaus."

SPMS is considered as a severe form of RRMS; usually it follows after RRMS. It develops after severe neurodegeneration and neuron loss, for which the CNS cannot compensate [89]. This state has been characterized by axonal degeneration as compared to RRMS which is considered to be inflammatory demyelination. In almost 50% RRMS cases, the disease changes course to SPMS within a decade of disease onset. 90% of RRMS cases change course to SPMS after 20-25 years [90–93]. Figure 2.5 shows the progress pattern in SPMS.

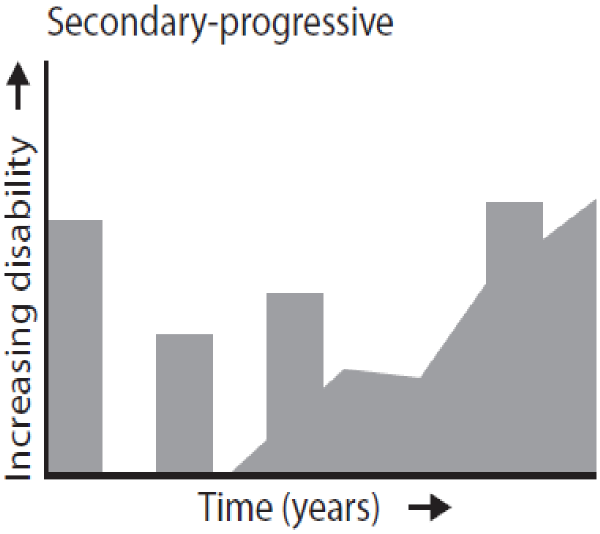


Figure 2.5: Disease progression pattern in SPMS. ~50% of people with RRMS will transition to SPMS within 15 years from diagnosis.

2.4.3 Primary progressive multiple sclerosis:

The standard description of Primary Progressive Multiple Sclerosis (PPMS) defined by clinicians [87] is as follows:

PPMS: "Disease progression from onset with occasional plateaus and temporary minor improvements allowed."

PPMS is characterized by progression without remission and relapse. It happens in almost 10-15% of the total patients. Unlike other forms of MS, there are no drugs available to cope with this subtype of MS. In this form of the disease, disability progresses without remission or improvement of the condition. This is the only form of MS in which there are no relapses. The age of its onset is also late as compared to RRMS (39 vs. 29 years). Another interesting fact is that the incidence of PPMS is almost equal in males and females [94].This suggests an etiology that is different from other forms of MS. Figure 2.6 shows the progress pattern in PPMS.

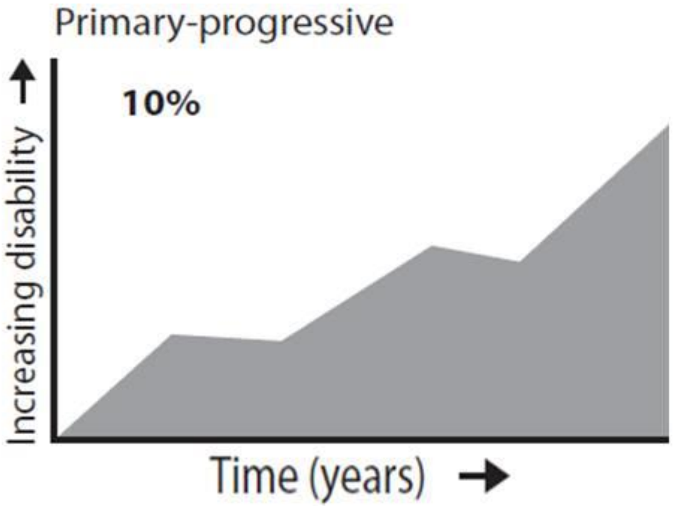


Figure 2.6: Disease progression pattern in PPMS, number of percentage shows the frequency of disease subtype.

2.4.4 Progressive relapsing multiple sclerosis:

The standard description of Progressive Relapsing Multiple Sclerosis (PRMS) defined by clinicians [87] is as follows:

PRMS: “progressive disease from onset, with clear acute relapses, with or without full recovery; periods between relapses characterized by continuing progression” [87].

PRMS is the least frequently occurring subtype of MS; it occurs in ~5% of the MS patients. Its distinct feature is disease worsening and progression with relapses. Figure 2.7 shows the progress pattern in PRMS.

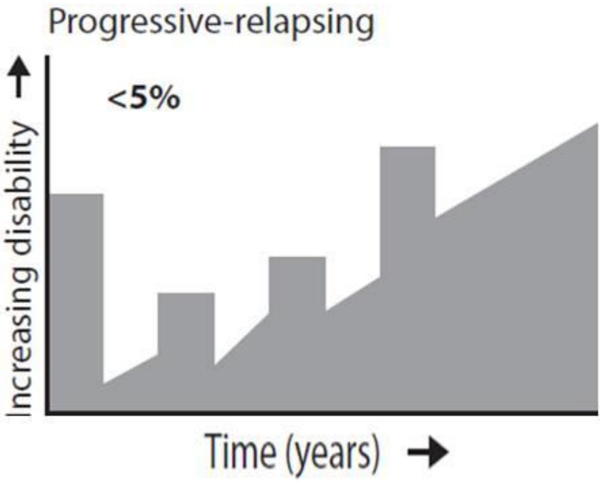


Figure 2.7: Disease progression pattern in PRMS, number of percentage shows the frequency of disease subtype.

Despite the variety of different disease subtypes with different phenotypes associated and the extensive study thereof, an underlying mechanism for any variation of MS has not been found. Treatment options for PPMS are therefore not available. The existence of different phenotypes implies that there are different underlying biomarkers, mechanisms, and different sets of molecules involved. Though scientists are aiming to personalize medicines based on the personalized genotypes of individuals, it is interesting to know that pathways involved in certain diseases subtypes are yet to be known. In next section we will discuss different types of biomarkers and MS biomarkers.

2.5 Biomarkers of MS:

A biomarker or biological marker is a measurable marker which indicates a biological state or condition. A classic example of a biomarker has been given as a laboratory parameter which can be used to help clinicians diagnose a disease and select appropriate treatment. The formal definition of a biomarker is as follows:

"A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention" [95]. Biomarkers are needed to distinguish different conditions in a biological system. Of the different types of biomarkers, the following four are the most commonly used.

A **Prognostic biomarker** helps to predict the disease course and its progression. It is a marker whose alteration or amount is measured or exists before the clinical end points of the disease outcome.

A **Predictive biomarker** is also known as a response biomarker. It helps clinicians to propose and prescribe a therapeutic course for the disease. By definition, it is a marker whose alteration or amount is measured or exists before the clinical end points of the treatment response. It is a marker whose alteration or amount predicts the response of a patient group to a treatment.

A **Diagnostic biomarker** helps to distinguish between a healthy state and a diseased one. It is a marker that is altered, or whose amount is modulated between the healthy and the non-treated disease states or between responders and non-responders (within the same disease). It illustrates the altered state between the diseased patients and healthy controls.

A **Pharmacodynamic biomarker** measures the treatment effects of a drug to help prescribe the right dosage. It is a marker whose alteration or the amount of modulation results from the effect of the studied compound/drug.

Like any other neurodegenerative disease biomarkers, MS biomarkers play a crucial role in diagnosing the disease, predicting the disease course, segregating the patient based on therapeutic response and optimizing the dosage response. Biomarker discovery is an important factor in MS, as the disease is heterogeneous in its clinical manifestations; conditions vary from patient to patient. Different kinds of biomarkers have been discovered and can be categorized, for example, by type (DNA, proteins and mRNA etc) or by the processes they are involved in (demyelination, oxidative stress and remyelination). The genetic biomarker HLA is one of the biomarkers widely accepted and associated with MS. It has been shown that two different types of HLA molecules exert different functions in MS. While soluble HLA-I has a role in neurological disorders and its presence in cerebrospinal fluid is linked with an increase in disease activity, soluble HLA-G is found to play a role in remission of disease [96–98]. One of the best known biomarkers for MS is the occurrence of Oligoclonal bands (OCBs) in cerebrospinal fluid, specifically increased level of IgG index [99]. OCBs are considered as a prognostic biomarker with higher specificity and sensitivity. It has been shown that OCBs are key indicators for the conversion of a clinically isolated syndrome into RRMS [100,101]. In addition, the presence of IgM OCBs has been associated with an aggressive disease course [102,103]. Besides molecular markers, there are many imaging biomarkers associated with MS which are considered as authentic prognostic biomarkers; for example, the existence of black holes, the volume of T1 ,T2, and gadolinium-enhancing lesions [99,104].

Category	**Biomarkers**
Diagnostic biomarkers	IgG oligoclonal bands; aquaporin-4 antibodies; heat-shock proteins
Predictive biomarkers	IgG and IgM oligoclonal bands; anti-MBP and anti-MOG antibodies; CHI3L1; Fetuin-A; TOB1; anti-EBNA1
Process-specific biomarkers	
1. Inflammation	Cytokines; chemokines; adhesion molecules; MMPs; osteopontin; sHLA-I and sHLA-II
2. Demyelination	MBP and degradation products; CNPase; 7-oxygenated steroids
3. Oxidative stress	NO and metabolites
4. Glial activation	S100b; GFAP
5. Remyelination/repair	NCAM; CNTF; BDNF; NGF; Nogo-A
6. Neuroaxonal damage	NSE; Nf and anti-Nf antibodies; tau; NAA

Table 2.3: Proposed molecular biomarkers associated with different phenotypes and disease mechanism, edited version of [105]. MOG: myelin oligodendrocyte glycoprotein; MBP: myelin basic protein; CHI3L1: chitinase 3-like 1; TOB1: transducer of ERBB2, 1; EBNA: nuclear antigen 1 of the Epstein–Barr virus; MMPs: matrix metalloproteinases; sHLA-I and sHLA-II: soluble HLA class I and II; CNPase: 2':3'-cyclic nucleotide 3'-phosphodiesterase; NO: nitric oxide; GFAP: glial fibrillary acidic protein; NCAM: neural cell adhesion molecule; CNTF: ciliary neurotrophic factor; BDNF: brain-derived neurotrophic factor; NGF: nerve growth factor; NSE: neuron-specific enolase; Nf: neurofilaments; NAA: N-acetyl aspartic acid.

Table 2.3 shows a few of the proposed molecular biomarkers with the associated disease phenotypes or specific processes e.g. inflammation or demyelination. This could be helpful to segregate and creating a unique profile of each patient based on the processes. In addition, monitoring certain processes (e.g. remyelination) after the administration of a drug would help health care professionals to formulate appropriate medication regimens. An up-to-date and comprehensive list of molecular biomarkers is discussed later (Table 2.5).

Besides molecular biomarkers, there are also anatomical biomarkers as well as techniques which allow us to quantify the phenotypes and identify the disease severity. These include number and volume of different lesions in different regions of the brain. Table 2.4 provides a list of techniques and potential imaging biomarkers associated with them.

Technique	**Markers**
T1	Number and volume of T1 lesions; presence of black holes
T2	Number and volume of T2 lesions
Gd	Number and volume of Gd-enhancing lesions
Brain volume	Brain parenchymal fraction; grey matter volume; white matter volume; cervical spinal cord volume; regional volumes
Magnetization transfer	Magnetization transfer ratio
MR spectroscopy	NAA, glutamate, glutamine, GABA, choline, creatinine, myoinositol, ascorbic acid
Diffusion MRI	Mean diffusivity; diffusion tensor
Functional MRI	Regional activation
Fractal dimension	White matter FD; gray matter FD
Optical coherence tomography	Thickness of the RNFL; macular volume

Table 2.4: Different techniques and imaging biomarkers for MS prognosis, edited version of [105]. MRI: Magnetic resonance imaging; Gd: gadolinium; NAA: N-acetyl aspartic acid; GABA: Gama Aminobutyric acid; RNFL: retinal nerve fiber layer.

Villoslada et al., [106] recently published a comprehensive list of MS associated molecular biomarkers with their types and status (possible, known or exploratory). The types of biomarkers include HLA, Activation markers, Adhesion molecules, Antibodies, Antigens, Cell phenotypes, Chemokines, Complement, Cytokines, Genes, Lipids, Metabolites, mRNA, Oligoclonal bands, Proteins and Viruses. The categorization is not strict but the table provides a summary of molecules for possible candidates for better diagnostics, drug discovery, and treatment of the disease.

Type	Examples	Status[1]
	DNA	
HLA	HLA-DRB1 (1501, 1503, 0801, 0301, 0401, 1401), DRB5, HLA-DQA, HLA-DQB (0603), HLA-C	Possible
Gene[2]	IL7R, IL2RA, CLEC16A, CD68, CD226, RPL5, DBC1, ALK, FAM69A, TYK2, CD6, IRF8, TNFRSF1A, SCIN, IL12A, MPHOSPH9, RGS1, KIF21B, TMEM39A	Possible
Genes[3]	ADAMTS14, AGER, ALS2, ALOX5, BANK, CD226, CCDC97, CYP2S1, CTLA4, FAM5A, LECAM2, GCCR, GSK3B, GPC5, AFGF, E1BAP5, ITGA4, ICAM1, IRF1, IFNGR1, IFNGR2, IL10, IL12, IL13, IL2RA, IL23R, IL3, IL4, IL4R, IL5, IL6, IL7, IL7R, IL9, CMT2A, GLOD1, PTPRC, FDC, LFA3, MMP7, MMP9, TIMP3, MICB, MAPT, SLC25A8, MBP, MAG, MPO, CMT1F, NPAS3, NPTXR, NT3, CARD15, OPN, CMT1A, PAI1, PECAM1, PLA2G7, PRR2, POU2AF1, GGF2, NKNA, JAG1, PKCA, HIP, PON1, STAT1, FLJ22950, LAP18, MMP3, SOD1, SYN3, PLAT, TCF7, TGFB1, TGFB2, TNFA, NGFR, GITR, TNFR2, TNFRSF5, 4-1BB, AXL, VEGF, VAMP	Exploratory
Genes[4]	CASP3, TRAIL, FLIP, COL25, GPC5, HAPLN1, CAST, STAT1, IFNAR1, IFNAR2, MX1, IFNG, IL10, GRIA3, CIT, ADAR, ZFAT, STARD13, ZFHX4, FADS1, MARCKS, IRF2, IRF4, IL4R, CASP10, CASP7, IL8, IFIT3, RASGEF1B, IFIT1, OASL, IFI44, IFIT2, HLA-DRB1*1501, TCRB, CTSS	Exploratory
mRNA	PDGFRA, BAX, BCL2, APAF1, API1, CASP1, CASP2, CASP6, CASP8, CASP10, P53, COL3A1, DOCK10, ADAM17, EGR2, EPHX2, EAAT1, G3PD, C11, HBB, HAVCR, IFI6, IFITM1, IFITM3, IFNAR1, IFNAR2, ISG15, MX1, G10P1, G10P2, IL1B, IL1A, IL10, IL12, IL4, IL5, CLEC5B, LY6E, LT, LAPTM5, MIF, MBP, MYD88, SIR2L1, NOTCH2, FLJ00340, EBP-1, RIP15, PRDX5, PLSCR1, PSEN2, PDCD2, PDCD4, PARK7, JAG1, PKB, RSAD2, EB9, HIP, NOGO, STK17A, TLR4, TLR6, NFKB3, TGFB1, TRIB1, TNFA, TRAIL, TNFSF12, APRIL, FASL, TNFRSF12A, UBE4B, XIAPAF1, RASGEF1B, OASL, MARKS	Exploratory
	Proteins	
Oligoclonal bands	IgG index, IgG OCB, IgM OCB, light chains	Known
Antibodies	Anti-MBP, anti-MOG, anti-GalC, anti-PLP, anti-OSP, anti-CNPase, anti-transaldolase, anti-proteasome, anti-β-arrestin, anti-Gangliosides, anti-CRYAB, anti-HSP60, anti-HSP70, anti-HSP90, anti-ATP2C1, anti-KIAA1279, anti-PACSIN2, anti-SPAG16, anti-hnRNP B1, anti-Alu repeats, anti-NG2, anti-phosphatidylcholine, anti-NF, anti-NogoA, anti-tubulin, anti-enolase, anti-glycan, anti-triosephosphate isomerase (TPI), anti-GAPDH	Exploratory
	Anti-AQP4[5]	Known
	Neutralizing antibodies of interferon β or Natalizumab	Possible
Cytokines	IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-10, IL-12, IL-13, IL-15, IL-17A, IL-18, IL-23, TNF-α, TGF-β, interferon β, interferon γ	Exploratory
Chemokines	CCR2, CCR5, CCR7, CCL1, CCL2, CCL3, CCL4, CCL5, CCL8, CCL17, CCL21, CCL22, CXCR3, CXCR4, CXCL5, CXCL10, CXCL12, CXCL13	Exploratory
Complement	C3, C3d, C4, C7 neoC9	Exploratory
Adhesion molecules	ICAM-1, VCAM-I, E-selectin, L-selectin, LFA-1, VLA-4	Exploratory
Activation markers	CD25, CD40, CD80, CD86, CD26, CD30, OX40, Fas, TRAIL, OPN, CD127, CD45, CD47, CD16, CD279, CD163, T-bet, CD1d, CD266, GITR, TNFR2	Exploratory
Other	αβ-Crystallin, neurofilaments (light-chain), tau, actin, tubulin, 14-3-3, neuronal enolase	Possible
	Nogo-A, Lingo, ALDH, α1B glycoprotein, α2-HS-glycoprotein, α-synuclein, Aβ, ANX1-5, ApoA (I, IV, B, D), API1, βADRBK1, Arrestin, beta 1, beta-End, NGF, BDNF, CNTF, BRCA1, CRP, CB2, CD276, CD44, chitotriosidase-1, chromogranin A, clusterin, contactin1, cystatin C, CD26, Mac-2 BP, gelsolin, GFAP, haptoglobin, iNOS, IGFBP-3, interferon α, interferon γ, MxA, IL-1ra, kallikrein-1, kallikrein-6, Manan-binding lectin serine protease-1, MMP-9, TIMP-3, MICB, MBP, MAG, NT3, OLIG2, P2X7R, PDGFB, PD-L1, PD-L2, IGFBP3R, COX-2, DJ-1, PACSIN2, protein C inhibitor, S100A, S100B, RBP4, secretogranin I, transferrin, serum paraoxonase/arylesterase 1, Stat-1, SCN2A, Sox-9, Sox-10, SPAG16, MMP-3, SOD1, tetranectin, tPA, transferrin receptor, TGF-β, peripheral benzodiazepine receptor, transthyretin, TNFSF12, tissue factor, Fas, vitamin D-binding protein, VDAC1, AZGP1	Exploratory

Type	Examples	Status[1]
Metabolites	Folic acid, homocysteine, prostaglandin E2, vitamin D, vitamin B12 , vitamin B6, hydroxyindoleacetic acid, iron, malonaldehyde, N-acetylaspartate, neopterin, nitrates, orosomucoid, sorbitol, thiobarbituric acid reactive species, cholesterol, 24S-hydroxycholesterol	Exploratory
Lipids	Galactocerebroside, gangliosides, sphingolipids, phosphatidyl-serine, oxidized cholesterol derivatives	Exploratory
Antigens	MOG, MBP, PLP, β-arrestin, contactin 2,	Exploratory
	AQP4	Known
Cell phenotypes	Treg (Foxp3+, Tr1, CD8reg)	Possible
	Breg, NK cells, CD4, CD8, B cells (CD5^{+}), macrophages, DC (myeloid and plasmacytoid)	Exploratory
Viruses	EBV, HHV-6, MSRV, VZV	Exploratory

[1] Status shows different types of biomarkers taken from Integrity^{TM}
[2] Genes found and validated in GWAS
[3] Genes marked as biomarker in Integrity^{TM}
[4] Genes identified in pharmacogenomics studies in Interferon-beta therapy

Table 2.5: A comprehensive list of MS biomarkers, their type and status, edited version of [106].

2.6 Therapies of multiple sclerosis:

Few therapeutic options are currently available to MS patients, and none cure or eliminate the disease completely. The available pharmacology regimens either only suppress the progression of the disease by reducing the symptoms, or bring about the recovery phase in certain disease subtypes. Further, PPMS does not have any approved drug thus far [107]. Most of the drugs have an influence on immunosuppression, thus preventing an immune system attack. Different types of drugs are prescribed based on the clinical manifestations of the disease. Corticosteroids are given to reduce attacks, e.g. prednisone or intravenous methylprednisolone, which reduce nerve inflammation. Beta interferon and Glatiramer acetate have been shown to reduce the frequency of relapses and stop autoimmune attacks on the myelin sheath, respectively. Both of the drugs are used as first line therapy and are the first choice of treatment by clinicians after diagnosing the disease. Another first line drug is Fingolimod, which has been recently approved and became the first oral drug for MS. Fingolimod inhibits lymphocyte emigration from lymphoid organs. Natalizumab, a strong antibody which targets VLA-4, is often prescribed as a second line therapeutic agent. It has a very high efficacy, but is also associated with side effects, such as progressive multifocal leukoencephalopathy.

The sequential line of therapies suggests that the drugs being prescribed as first line therapy are targeting the molecules and pathways involved in the early stage of MS. This provides intrinsic knowledge about the order of disease progression, e.g. molecules affected by first line therapy would activate earlier than molecules affected by second line therapy. Following the path of modeling molecules involved in different lines of therapies would unravel a chronological model of MS. Table 2.6 shows a list of drugs segregated by their respective lines of therapies (1st line or 2nd line) and their mechanism of action. By looking at the mechanism of action, one could argue that the progression of the disease follows the same pattern.

Group	**Therapy**	**Proposed mechanism of action**	**Comments**
First line	Interferon-ß	Complex (i.e., inhibition of BBB transmigration and Th17-cells, promotion of regulatory lineages)	Moderate efficacy, Good side effect profile, Neutralizing antibodies
	Glatiramer acetate	Complex (promotion of regulatory lineages)	Moderate efficacy, Good side effect profile
	Dimethyl fumarate	Not clear, Immunomodulation	Moderate efficacy, Oral agent
	Teriflunomide	Pyrimidine synthesis inhibitor, Impairs T-cell activation	Moderate efficacy, Risk of hepatotoxicity, Oral agent
	Fingolimod	S1P-receptor modulator, Inhibits lymphocyte emigration from lymphoid organs	High efficacy, Risk of infections and cardiac side effects, Oral agent
Second line	Natalizumab	Anti-CD49d antibody, Inhibits BBB transmigration	High efficacy, Risk of progressive multifocal leukoencephalopathy
	Mitoxantrone	Immunosuppressive agent, Induces lymphopenia	High efficacy, Risk of cardiomyopathy and secondary leukemia
	Azathioprine	Immunosuppressive agent, Induces lymphopenia	Moderate efficacy, Risk of malignancies, Oral agent
Promising agents	Rituximab /Ocrelizumab	Anti-CD20 antibody, Depletes B-cells	High efficacy, Side effect profile unclear
	Daclizumab	Anti-CD25 antibody, Induces regulatory NK-cells	High efficacy, Side effect profile unclear
	Alemtuzumab	Anti-CD52 antibody, Depletes lymphocytes	High efficacy, High risk of B-cell mediated autoimmunity
	Laquinimod	Not clear, Immunomodulation	Oral agent

Table 2.6: Different lines of therapies and their proposed mechanisms of action, edited version of [35].

A recent review demonstrates that different approaches are being taken to treat MS by targeting multiple sets of molecules based on the vicinity of the drugs' actions [108]. Figure 2.8 shows various drugs acting on different compartments. Rituximab, Alemtuzumab, Daclizumab and Natalizumab are monoclonal antibodies (denoted by the suffix –mab) which interact with different CD molecules, NK cells, and VLA-4 in the periphery of blood-brain barrier. Cladribine and Teriflunomide are therapies which prevent new T-cell formation. Fingolimod (FTY720) interacts with sphingosine-1-phosphate receptor (S1P-R) in the CNS or in BBB periphery (Figure 2.8). Fingolimod is also able to permeate through the BBB. In the CNS, Laquinimod and Fumaric Acid play a role in the reduction of disease severity via the interaction with T-helper cells. Laquinimond inhibits TH17 while Fumaric Acid interacts with TH2 cells.

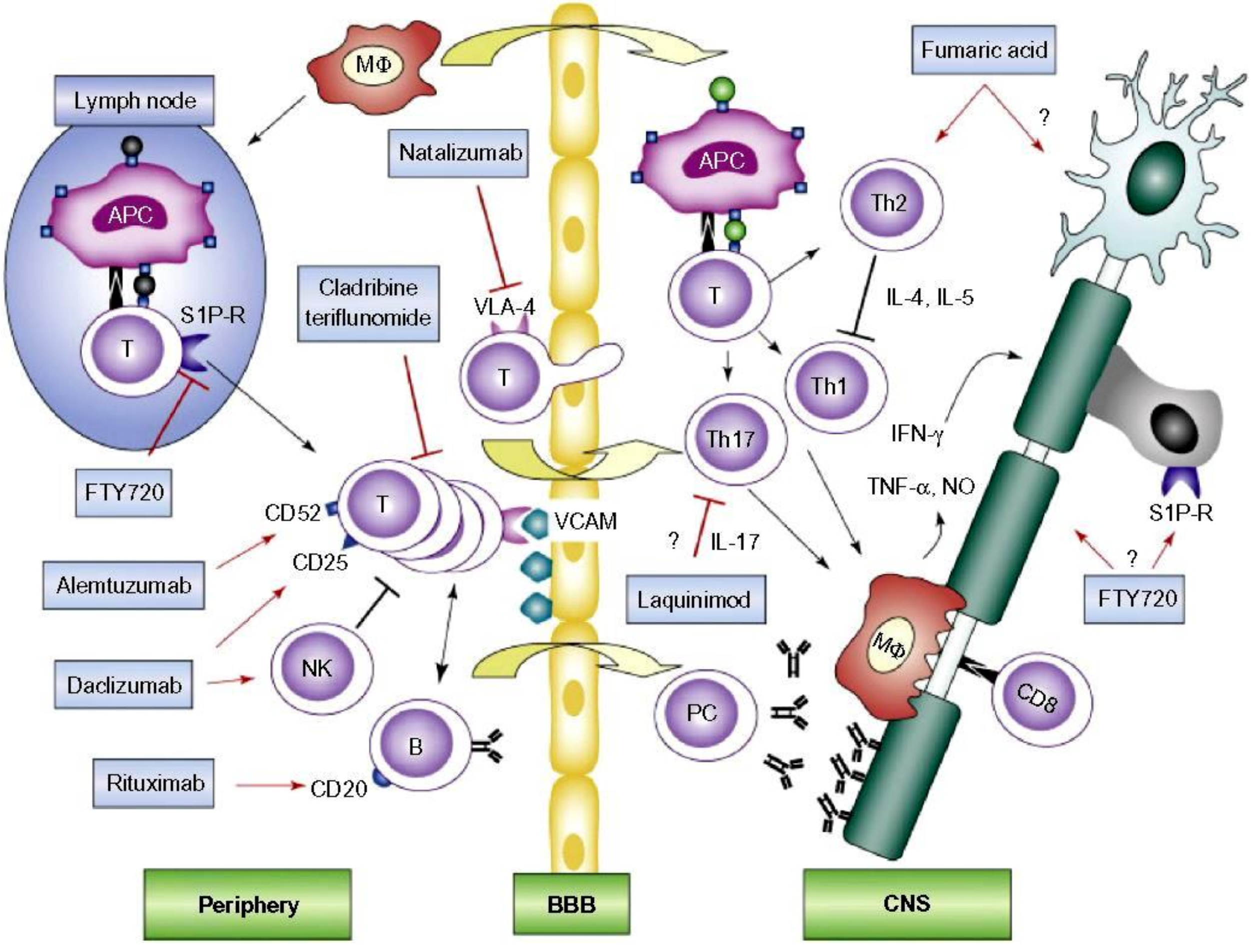


Figure 2.8: Drug actions on the different molecules associated with MS, taken from [108]. The blue square boxes show the name of drugs. The round shapes represent the different types of immune cells. The green boxes represent the sites where these interactions are taken place.

Despite the knowledge and variety of drug treatments available, the struggle to stop and reverse neural degeneration in MS continues. Although the drugs discussed above provide relief of symptoms, it is only short term relief, and a complete cure is still beyond reach. In addition to the lack of cure, no treatment exists which slows an aggressive form of the disease such as PPMS. One reason a cure has not yet been found may be that researchers currently focus primarily on various single aspects of the disease. MS is multi-factorial and systematic; one might infer that a cure can be found through a multi-factorial and systemic approach. Systematic approaches can help find a cure for diverse polygenic diseases, as demonstrated in the case of breast cancer, in which an experimental systemic therapy was found to reduce metastatic extravasation [109].

3 Chapter 3: Systems Biological approaches to model MS

3.1 Introduction to Systems Biology:

Systems biology is an emerging field which enables scientists, particularly biologists, to look at the whole picture of a specific biological phenomenon. One of the rationales behind systems biology is the notion that, “The whole is greater than the sum of its parts”. There are many definitions of systems biology [110]; one of the earliest and simplest definitions is from Kitano: “A systems biology approach requires the integration of experimental and computational research to understand complex biological systems” [111]. Systems biology is crucial, since complex diseases are polygenic and therefore require a better understanding of the non-linear interactions among the molecules involved in their pathogenesis. Drugs against complex diseases can only be effective if networks, rather than individual molecules, are targeted. Systems biology is also known as integrative biology, as depicted in figure 3.1. Different approaches have been integrated to solve biology problems, e.g. information sciences, system sciences and biological sciences. The integrative approach allows researchers to connect multiple layers of datasets in a multi-dimensional paradigm for complex mechanism modeling.

Systems biology aims to investigate biological systems by capturing and integrating global biological datasets from different hierarchical organizations, in order to visualize emergent properties. The term emergent properties, also known as collective properties, is used to describe the intrinsic properties of a system which cannot be predicted by studying individual components of the system [112]. Systems biology uses a holistic approach, utilizing all technological advancements to obtain the minute details and features of each component of a system. Computational tools, such as data mining, help to reveal the knowledge buried under piles of data. System sciences help to build models, at different resolutions by using multi-level datasets, which describe various complex biological phenomena and provide analytical insights. The availability and interoperability of publicly accessible datasets of different types and levels such as genome sequences, Protein-Protein interactions, high throughput and functionally annotated databases made it easier to interconnect different molecules with their contextual information in-silico, thus facilitating simulation and prediction. In summary, systems biology makes use of approaches taken from different domains, as well as addresses challenging tasks such as drug discovery, polypharmacology, pharmacovigilance, and personalized medicine [113].

The following four areas are considered to play a key role in systems biology [114]:

(1) Genomics and other molecular biology research

(2) Computational and bioinformatics tools, e.g. modeling and simulation software

(3) Analysis and dynamics of the system

(4) Technological advancement for high precision measurements

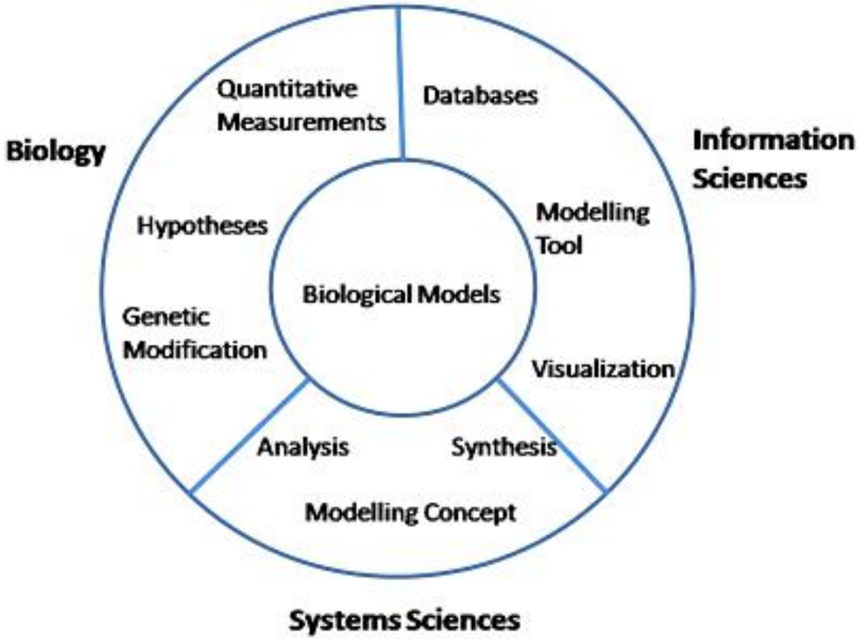


Figure 3.1: A descriptive diagram of Systems Biology, edited version of [114].

As discussed above, the ultimate applications of systems biology are to foster drug discovery and to unravel the molecular mechanisms behind complex processes. Even though the application of modern molecular biology techniques significantly advanced the way drugs are being designed, the struggle to find a cure for many diseases continues. For example, the discovery of new isolation techniques allowed researchers to identify the molecules responsible for disease conditions, observe molecular behavior after treatment with certain compounds, and discover molecular interactions. More recently, the post-genomic era brought a large amount of new data such as genes, gene products [115], and protein connectivity maps [116–118]. Remarkable progress has been made in the ability to acquire this data. In order to cure complex diseases, the interpretation of such a large amount of data requires a systematic approach. Biological systems should be considered in their entirety.

In the recent past, it was discovered that most of the common diseases such as cancer, neurodegeneration and cardiovascular diseases are the outcome of many different molecules' dysfunction [119,120]. Additionally, researchers discovered that the human genome has many more levels of organization and regulation than previously anticipated; thus a systematic level study of organization and interpretation is essential. This drives researchers to look not only at any individual molecule, but rather to study them in relation to their often complex interactions. One common approach in systems biology is to develop a molecular interaction map, analyze the pathways, model molecular systems to find out the key molecules involved in the disease state, and then use this knowledge for therapeutic intervention [121]. Systems biology allows researchers to look at the complexity of biology as a whole and model knowledge with different layers of complexity, as well as to make predictions based on the state of the system. With this kind of systematic approach, a study [122] demonstrated that in response to a drug, approximately 1000 different molecules of a H1299 lung carcinoma cell behaved differently than they did before treatment. This shows that the perturbation caused by drugs is not limited to a few molecules, but rather affects large networks. This demonstrates that a better understanding of drugs' actions is only possible through analyzing the whole system. Table 3.1 shows the comparison of Traditional drug discovery approach vs. Systems biology approach.

Traditional drug discovery approach	**Systems biology approach**
Reductionist approach	Holistic approach
Inhibition of single drug target	Inhibition of one or more key targets at converging point in disease pathway
Non-account for organism's compensatory mechanism	Account for organism's compensatory mechanism
High risk in animal model to clinical translation	Animal model to clinical translation guided by biomarker(s)
High risk in clinical study relying solely on efficacy endpoints	Mitigated risk in clinical study as biomarker reduction used as early decision endpoint
Nonconforming to personalized medicine	Facilitating personalized medicine

Table 3.1: Comparison of Traditional drug discovery approach vs. Systems biology approach. Edited version of [123].

In addition to the discovery of new drugs, the improved understanding of entire complex systems could be used to prevent diseases and improve healthcare by allowing scientists to focus on the key molecules involved. Another example of this systematic approach is an experimental therapy against breast cancer, in which EREG, COX2, MMP1, and MMP2 were found to be key genes. Inhibitory action against them resulted in a spectacular reduction of metastatic extravasation [109]. With this evidence that one often needs an alternative such as systems biology in order to understand and affect complex diseases, one might expect that systems biology itself offers a range of options to suit various investigative or treatment purposes. Two well-known systems biology approaches, bottom-up and top-down, will be discussed in the following sections.

3.2 Bottom-up systems biology:

The bottom-up systems biology approach starts with fine granularity. The interactions of molecules are studied in order to unravel their functional attributes. Individual components of the systems are modeled to provide the mechanistic details, dynamics, and simulations closer to actual biological processes. The bottom-up approach can be started by a simple literature search about a molecular entity and exploring interactions. Enrichment of the interactions with specific experimental studies depends on the interests of the researcher and on data availability. For this approach, prior knowledge of most of the molecular interactions is required.

Different kinetic models, which represent and simulate signaling processes, e.g. the virtual heart model [124] and the silicon cell model [125,126], have been developed by using this approach. Besides models, many databases have been assembled by using the same approach e.g. KEGG [127] and EcoCyc [128] etc.

3.3 Top-down systems biology:

The top-down systems biology approach attempts to build in-silico models by processing high-throughput experimental studies. This approach helps to model the system on a high-level resolution, considering only the input and output behaviors of a model. The top-down approach is ideal for experimental methods that are data rich; it facilitates valuable insights from large datasets by modeling and inferences [129,130]. One of the advantages of this approach is that, often, prior knowledge is not needed to generate hypothesis; thus many different pathways can be analyzed simultaneously. Contrary to the bottom-up approach, the top-down approach is based on data and can be considered a data-driven approach. Large datasets are required of gene expression, proteomics, affinity assays, etc. to work with this approach. The starting point of this model building approach is to have a comprehensive dataset, often generated by –omics technologies (genomics, proteomics, metabolomics etc.). The dataset is then used to infer the mechanisms underneath, such as phenotypes or disease subtypes, by identifying specific pathways or modules and their components. One of the strategies for using this approach is the integration of various datasets to provide a dynamic state of the model. The top-down approach can be applied to different phases of system modeling e.g. the module discovery phase, the knowledge-based static modeling phase, or in various later phases. Bayesian modeling may be used to unravel the regulatory components. The top-down approach can also be used for time series models, e.g. disease progression or changes in physiological states. In addition, the approach can also be used to observe perturbed systems [131], e.g. to monitor a drug therapy and any global changes it might cause. Gene-set enrichment analysis, pathway mapping, disease classification, reverse causal reasoning, and network inference are few methods categorized under this approach [132]. Clustering techniques have also been used to infer novel relationships among groups of genes by using the same approach [133–135].

A few drawbacks of using the top-down approach are: limited opportunities to discover a unique and specific mechanism without intertwined interactions, findings which are often based on correlation rather than cause and effect, and difficulties in finding relevant supportive knowledge of a network once a correlation between genotype and phenotype has been revealed. While the bottom-up approach does not produce results with these complications, it is nevertheless limited in that most results are obtained using transgenic animals rather than humans. Thus, the accuracy of this knowledge in the context of human disease is questionable [136]. Figure 3.2 shows the different perspectives of the top-down and the bottom-up Systems biology approaches.

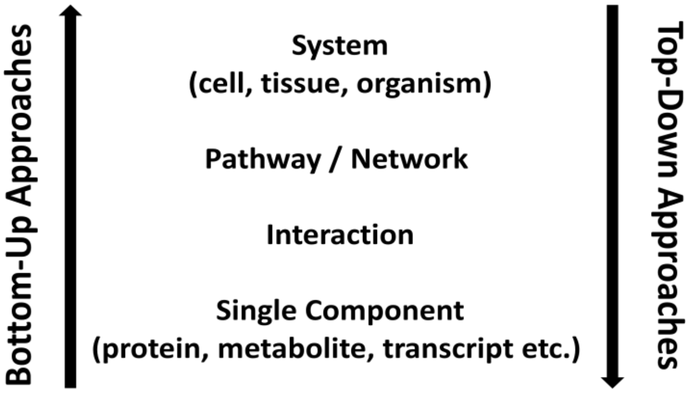


Figure 3.2: Difference between the two Systems biology approaches. Top-down approach starts from System and then goes to single component, while Bottom-up approach starts from single component and then goes to System.

Bottom-up approach has been used to model many molecular interaction maps of diseases with the initial seed entities, and then to expand the network with other relevant knowledge. These networks have been developed to discover potential drug candidates for diseases such as rheumatoid arthritis [137], Alzheimer's disease [138] and hepatitis B [117]. In addition to the bottom-up approach, the top-down approach has also been widely used on high throughput datasets; for example, to elucidate the mechanism behind neurotoxicity in Alzheimer disease [140]. Since both approaches have their drawbacks, an integrated approach would be the answer to the challenges and may accelerate the discovery of new drug targets due to the unique strengths of each of the approaches. In the following section various computational techniques and approaches have been discussed to model the disease.

3.4 Text mining and systems biology:

Text mining is a computational technique used to discover new and unknown information by processing written resources [141]. There are many processing steps, including information retrieval, information extraction, and natural language processing. Text mining is one of the crucial techniques in systems biology toolkit. It facilitates the rapid processing of millions of documents, as well as the extraction of important information from them. With this capability, text mining could provide a solution to help the scientific research community cope with the rapidly increasing numbers of scientific publications. An estimated 500,000 new citations are added each year to PubMed, which already contains approximately 24 million citations [142].

Text mining has been used extensively to identify disease-related molecules, such as genes and proteins, and to understand their functions. Many databases use text mining tools to process entire contents of PubMed and extract information from the available scientific text for novel discoveries. In addition to the increasing number of publications, the availability of databases is also growing in number and variety; for example, microarray, genome-wide association study (GWAS), and pathway databases. The table 3.2 below shows a list of some of the databases and tools:

Text/Structural databases	
PubMed Central	http://www.pubmedcentral.nih.gov
HighWire Press	http://highwire.stanford.edu
E-Biosci	http://www.e-biosci.org
PubMed	http://www.ncbi.nlm.nih.gov/pubmed
UniProt	http://www.uniprot.org
InterPro	http://www.ebi.ac.uk/interpro/
Text Mining Tools	
Google Scholar	http://scholar.google.com
GoPubMed	http://www.gopubmed.org
Textpresso	http://www.textpresso.org
BioRAT	http://bioinf.cs.ud.ac.uk/biorat
ABNER	http://pages.cs.wisc.edu/~bsettles/abner
iHOP	http://www.ihop-net.org/UniPub/iHOP
GeneWays	http://geneways.genomecenter.columbia.edu
Microarray databases	
SMD	http://genome-www5.stanford.edu
Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/geo
Oncomine	http://www.oncomine.org
CGAP database	http://cgap.nci.nih.gov
caArray	http://array.nci.nih.gov/caarray
Gene Expression Atlas	http://symatlas.gnf.org
Clustering Platform	
GenePattern	http://www.broad.mit.edu/cancer/software/genepattern
GeneCluster 2	http://www.broad.mit.edu/cancer/software/genecluster2/gc2.html
ArrayMiner	http://www.optimaldesign.com/ArrayMiner/ArrayMiner.htm
Supervised Analysis Platform	
SAM	http://www.stats.stanford.edu/~tibs/SAM
Pathway and interactome databases	
KEGG	http://www.genome.jp/kegg
UniHI	http://theoderich.fb3.mdc-berlin.de:8080/unihi/home
PathwayExplorer	http://pathwayexplorer.genome.tugraz.at
GenMAPP	http://www.genmapp.org
Pathguide	http://www.pathguide.org

Table 3.2: Some of the popular websites for text mining, microarrays, pathway databases, and associated tools, taken from [143].

3.5 Systemic modeling approaches and their advantages:

Models are essential to the understanding, prediction, and control of the system; in biology, all three are needed for drug discovery. Models have been always important in life science research, whether they are living organism models or computational models. Modeling biological systems is a major application of systems biology. The development of models which represent and describe a biological system is crucial for simulating a physiological state in-silico. Systems biology provides a framework to integrate multi-dimensional data, thus facilitating the generation of models with complex attributes. Integration capabilities of the framework are required in order to have a holistic approach, which facilitates the understanding of the structures and dynamics of intercellular and intracellular interactions, as well as understanding cell functionality. Different granular layers of interactions can be modeled to understand the functionality of different modular components, with respect to their compartments.

Modeling and simulation made it possible to predict the output of a variety of scenarios without even conducting an experiment. Figuring out the interactions between macromolecules, cells, tissues, and their respective dynamics is always a challenging task for researchers. Mathematical formalization has been used to study biological systems [144–146]. In addition, system level modeling helps to modularize different components and dissect large components into individual elements, in regards to their underlying dynamic interactions. Mathematical modeling has been used widely at different abstract levels and it records the behavior of individual molecules with kinetic details. Many modeling methods e.g. ordinary differential equation, stochastic modeling, flux balance analysis, and metabolic control analysis have been used to model biological systems [147].

Ordinary differential equation (ODE) models usually require a large number of kinetic parameters; generally, the kinetic data is either limited or unknown. This makes ODE models a poor choice for our work, as kinetic data was not available for most of the interactions. Stochasticity plays an important role in biological networks [148], as it includes randomness in certain events which occur spontaneously in biological systems, e.g. switching from non-lytic to lytic mode of different phages and diffusion of molecules across a membrane. We used stochasticity partially in our agent-based modeling work (discussed in later chapters). Flux balance analysis (FBA) is mostly used to quantify the metabolic state of a cell by means of constructing metabolic networks and modeling metabolic pathways. Metabolic control analysis (MCA) is a method for analyzing key reactions of metabolism and has also been used to identify cell cycle reactions. One of the drawbacks of the approach is that MCA can only be performed if models are defined in terms of reaction stages. On the other hand, the higher level of abstraction in equations provides a generalization of the system phenomena.

The aforementioned modeling methods could help us to accomplish the following objectives:

- Optimizing strategies for certain drug treatments
- Identifying viable drug targets
- Understanding drug effects, pharmacodynamics, and pharmacokinetic properties
- Repurposing eligible drugs

In the following chapters, these aspects of modeling will be discussed briefly.

3.6 Systemic modeling and drug discovery:

Computational disease modeling (CDM) is an important method of systems biology by which pharmaceutical companies could avoid dried-up drug pipelines, cope with unmet medical needs, and further treatment options for patients. CDM is defined as the mathematical and computable modeling of a disease which can simulate the disease's state, progression, and dynamics. CDM has the potential to significantly reduce the timeline of drug development and the cost involved therein. Either of these hinders the progress of drug development. Developing a drug often takes 12-15 years and can cost up to $1.7 billion [149]. CDM could help in designing strategies to hit multiple targets and relevant pathways of a disease, while also predicting possible side effects of drugs [150–152]. A recent study demonstrated that CDM helped to optimize therapeutic strategies by identifying multi-factorial components of the system [117].It could also allow researchers to predict the outcome of combinatorial therapy, the best dosage possible, and the most favorable target. With computational disease models, one could simulate the disease condition and observe it in regards to the whole system. A success story of computer based modeling surfaced when the U.S Food and Drug Administration (FDA) approved the first ever computational model as a substitute for pre-clinical animal testing for new treatment of diabetes mellitus type 1 [154].

3.7 Qualitative modeling:

Qualitative modeling is a type of modeling which deals with the continuous aspects of a system. Sometimes it is also called qualitative reasoning [155]. This modeling technique is used when numeric details of system components are not known; instead, increments, decrements, highs, and lows are being used. The goal of this technique is to represent and reason the system computationally without quantitative values. This technique provides a possibility to model systems which are too complex to model otherwise, due to the different constraints and the descriptive nature of the system. Qualitative models aim to infer as much as possible from minimal observations and descriptions of the system, rather than from mathematical equations.

Qualitative models provide an abstract view of the system, as minute details are often overlooked; they specify objectives and their basic parameters, qualitative interrelationships, and underlying hypotheses. They transform an objective statement and a hypothesis into a conceptual model, which can be enriched with equations. These can then be transformed into quantitative models by the addition of mathematical equations, as shown in the case of Petri-net models [156]. Qualitative models can be presented in different forms but mostly are represented by diagrams. Diagrammatic models represent system entities as nodes and relations as edges, such as the models available in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [127].

3.8 Quantitative modeling:

Quantitative models require an extensive knowledge of their components and mathematical equations. Quantitative modeling is based on mathematical and statistical methods. This modeling approach is used to interpret numerical data and to develop mathematical models based on which simulations and hypotheses or predictions can be generated. It also makes use of measurements and the manipulation of variables to see the global effect of numerical value change.

The quantitative aspect of modeling often comes later than the qualitative aspect. This happens because modeling requires a system, which specifies the objects, their basic descriptors, the qualitative interrelationship, and any underlying hypotheses before it can enrich them with numerical values and equations. Quantitative models are very efficient for modeling system dynamics and providing accurate predictions, if sufficient data is available, which is unfortunately not the case with neurological disorders. However, there is large amount of textual data available and one can mine those to build some text based knowledge models. Molecular interactions maps are one of the many approaches to model disease after information retrieval.

3.9 Molecular interaction maps:

Molecular interaction maps (MIMs) are interaction maps of molecules that are involved in a biological function. These interaction maps may also be called interactomes [157], molecular interaction networks [157], protein-protein interactions [158], protein-DNA networks or gene regulatory networks [159], depending on their scope and components. They are based on graphic representations or diagrams of physical interactions among different molecules in an organism, either in a compartment or in a physiological state. As pointed out by Ideker et al. [160], graphic representation or diagrams can be "*a tremendous aid in thinking clearly about a model, in predicting possible experimental outcomes, and in conveying the model to others*". MIMs make it convenient to visualize all possible interactions of a molecule. Most proteins interact with many other molecules to form complexes and networks. Each human protein interacts with roughly 15 other molecules [161]. MIMs are developed from different sources; it could be experimental data, biochemical data, or published literature. MIMs are used to discover overall functionality and the regulatory processes of a biological system, as well as the pathophysiology of complex human diseases [162].In addition, they can be used to identify novel biomarkers, pathway crosstalk and molecular characterization of complex diseases [143]. In MIMs, each node represents a molecule (gene or protein) and each edge represents an interaction. Most MIMs are undirected, which means they do not show the flow of signaling between two molecules.

Exploring the disease mechanism has been considered an important application of MIMs, and many MIMs are developed [137,139,162–166] to reveal the mechanistic details of molecular machinery behind a disease state. Traditional disease maps do not distinguish cause and effect, as they lack the directionality of the events. These maps leave the impression that interacting nodes are somehow related, without explaining whether they are participants in a reaction or product. One approach to develop MIMs is based on information retrieval. It could be started by a simple search of PubMed and then ranking the molecules based on their association with the disease. Due to the increasing numbers of MIMs, a community-based portal Payao [31] has been launched to help scientific community with interactive knowledge sharing in systems biology languages.

3.10 Systems biology languages:

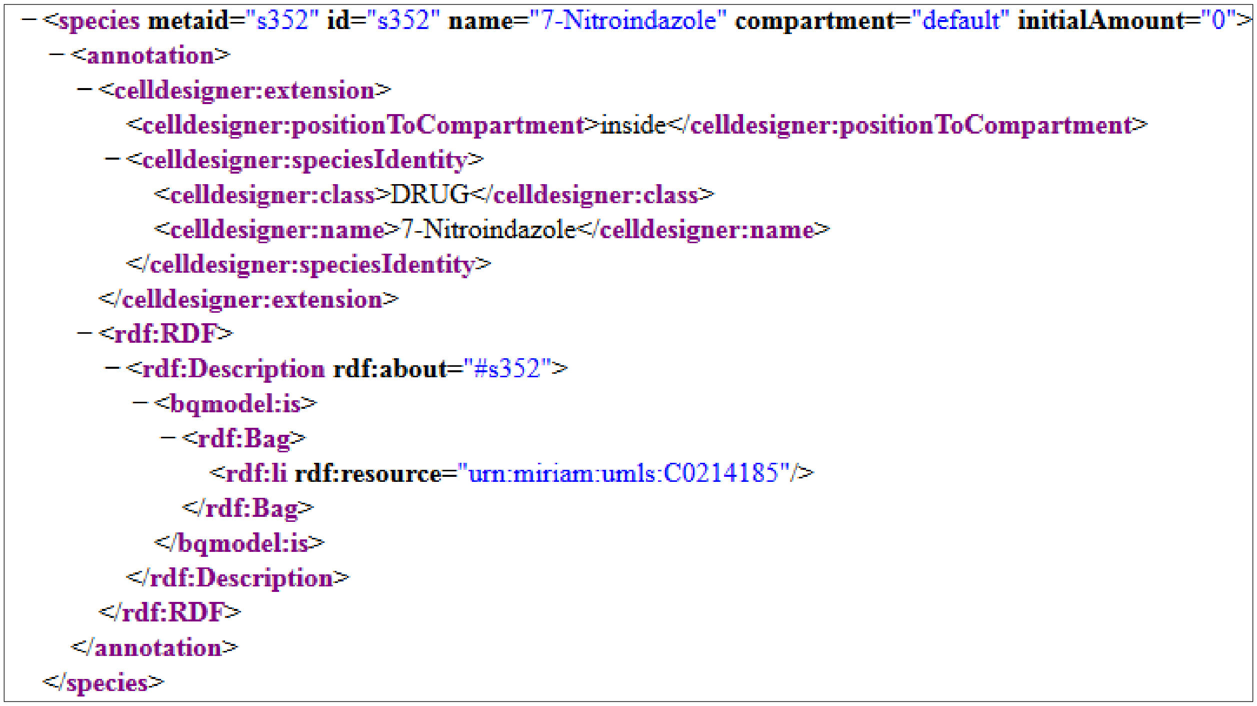
There are many languages which can represent biological data and facilitate the exchange of information from one platform to another. Many factors are involved in choosing the appropriate language for answering specific questions, based on the aim of the work. Factors include, but are not limited to, a variety of tools available to support the language, very expressive (in terms of biological reactions) or enrichment capabilities, and the preservation of knowledge when transformed into other formats. We were interested to model and simulate biological data with the possibility to use modeling knowledge in other platforms such as Mediawiki, thus systems biology markup language (SBML) was chosen. A plug-in has been developed which transforms SBML into wiki pages, thus allowing knowledge sharing beyond specific tools (discussed in later chapters). There are hundreds of tools available which extend the possibilities of SBML functionality; thus, SBML was the best choice for the current work on disease modeling. A brief description of various biological modeling languages, featuring their respective strengths and limitations, is given below.

3.10.1 Systems biology markup language (SBML):

Computational modeling in biology is no different from traditional computational modeling, except that models are developed from biological data. Like any other model the biological models are computable, can be simulated, and can be analyzed by mathematical methods. Many different representations of models are used for different purposes. Systems biology graphic notation (SBGN) is a graphical representation for biological processes [167]. To make models computable with enriched and dynamic biological systems, a quantifying format is required; (SBML) [168] satisfies this requirement.

SBML is a machine-readable format for model representation. This representation emphasizes the chronological order of biochemical events, such as molecular entities' transformation into complexes, and entities' reactions as involved in a biochemical network. SBML framework is suitable for representing models, including cell signaling pathways, metabolic pathways, biochemical reactions, and gene regulations.

SBML is based on Extensible Markup Language (XML) and is a representative format for computable biological models. It is free with open interchange format and extensive software support; to date, there are 280 software systems which provide support for SBML [169]. Many biological phenomena can be represented by SBML, such as metabolic networks, cell signaling pathways, regulatory networks, and infectious diseases. Since its inception, SBML has become the standard for systems biology models representation. It has enabled the exchange of models between different software tools, thereby streamlining and enhancing framework interoperability.



```
– <species metaid="s352" id="s352" name="7-Nitroindazole" compartment="default" initialAmount="0">
  – <annotation>
    – <celldesigner:extension>
        <celldesigner:positionToCompartment>inside</celldesigner:positionToCompartment>
      – <celldesigner:speciesIdentity>
          <celldesigner:class>DRUG</celldesigner:class>
          <celldesigner:name>7-Nitroindazole</celldesigner:name>
        </celldesigner:speciesIdentity>
      </celldesigner:extension>
    – <rdf:RDF>
      – <rdf:Description rdf:about="#s352">
        – <bqmodel:is>
          – <rdf:Bag>
              <rdf:li rdf:resource="urn:miriam:umls:C0214185"/>
            </rdf:Bag>
          </bqmodel:is>
        </rdf:Description>
      </rdf:RDF>
    </annotation>
  </species>
```

Figure 3.3: Example of encoded SBML Model. The knowledge is embedded with different Tags and in a structured format.

As it represents participant dependent, reaction-type processes, SBML is not specific to biochemical network modeling. The same formalism could also be used in other types of processes and in expressing different functions of the system. SBML also supports direct mathematical expression and formulas, extending its capability to merely represent biochemical reactions [170]. SBML has following two main purposes:

- Enabling the use of different software tools without reconstructing the models for each application-specific file format, thereby creating the ability to share the models among people using different applications.

- Extending the longevity of the models beyond the lifetime of the software used to develop them.

SBML's purpose is to serve as an exchange format used by different contemporary software tools, in order to communicate essential aspects of a model.

3.10.1.1 Main features of SBML:

SBML can develop models based on entities and reactions. An excellent feature of SBML models is that they can be deconstructed into constituent elements easily, thus making them interchangeable into various formats and then readable by different forms of tools being used.

Different software programs can read and transform SBML models into the required format for further processing. SBML allows model representation with rather arbitrary complexity. All constituents of the models are defined by using a specific type of data structure, and knowledge is embedded within various layers. The data structures determine the encoding layout in XML. Furthermore, each entity can have machine-readable metadata associated with it. These annotations can be used to encode reaction details among the entities in a model as well as to encode the external identifiers of the entities, e.g. UMLS CUIs and Drugbank ID. The BioModels database utilizes this feature by annotating each model and providing references of the associated resources, i.e. research articles, databases, pathways, etc. Annotations make a model more meaningful and semantically enriched for embedded knowledge sharing. Minimum Information Required in the Annotation of Models (MIRIAM) is a standardized set of metadata developed by the SBML community to facilitate the unified curation process of biological systems. SBML models the constituents of many components, such as function definitions, unit definitions, compartments, species, complexes and reactions etc. In next section System biology graphical notation will be discussed.

3.10.2 Systems biology graphical notation (SBGN):

Systems Biology Graphical Notation (SBGN) is a graphical representation format developed by modelers, biochemists, and computer scientists [171].The proposed usage of SBGN languages is to store, exchange, and reuse information about signaling pathways and gene regulatory networks. It has also been used in molecular interaction maps such as rheumatoid arthritis map [172]. SBGN has a simple syntax and easily understood semantics, thus it is one of the most widely used notations in systems biology (Figure 3.4).

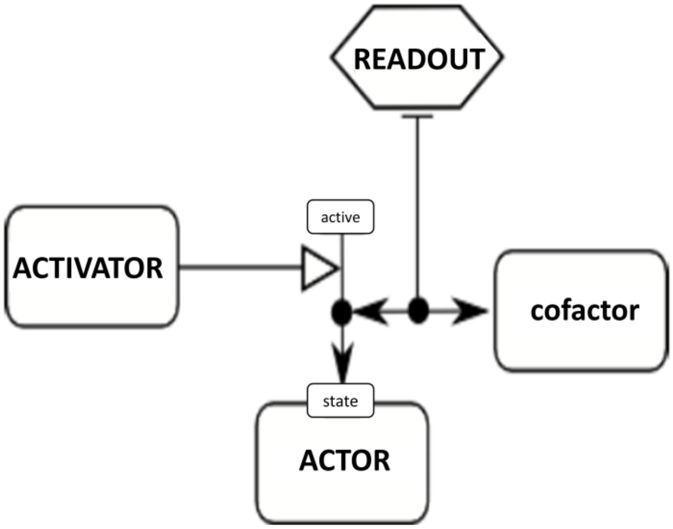


Figure 3.4: Example of SBGN Entity Relationship map [173].

3.10.3 Biological pathway exchange (BioPax):

Biological Pathway Exchange (BioPax) is a representative language for pathways. It is based on RDF/OWL, and as the name implies, it facilitates the exchange of pathway data. Exchange of pathway data is necessary to attain collective knowledge sets scattered throughout different databases and which may exist in incompatible formats. BioPax provides an easy way to accomplish different pathway data tasks such as gathering data, indexing data, data interpretation and data sharing. With the support of BioPax, thousands of pathways have been organized by millions of interactions found in many organisms, making them computable. Many online databases offer BioPax export, including Reactome [174], BioCyc [175], BioModels [170], Pathway Commons [176], and WikiPathways [177] are among many others. It is also supported by various tools such as Paxtools [178], Systems Biology Linker [179], ChiBE [180], BioPax validator [181] etc.

3.10.4 Biological expression language (BEL):

Biological expression language (BEL) is a relatively newly formed language; it represents scientific findings in a computable format. BEL has additional capabilities to capture contextual, causal, or correlative relationships. It also supports embedding observations and published evidence to provide a broad contextual knowledge within a model. The knowledge can be included during model development to facilitate qualitative modeling of biological processes.

BEL also supports automated reasoning methods such as reverse causal reasoning [182]. The enriched models developed in BEL are called Knowledge Assembly Models (KAM). Since BEL is a recently launched language, the numbers of applications are limited; thus a framework, which can process computable knowledge models, BEL Framework, has been designed. Like SBML and BioPax communities, a dynamic community is working together on the BEL framework to improve its functionality and enhance its capabilities. Figure 3.5 shows a simple BEL model with different reactions.

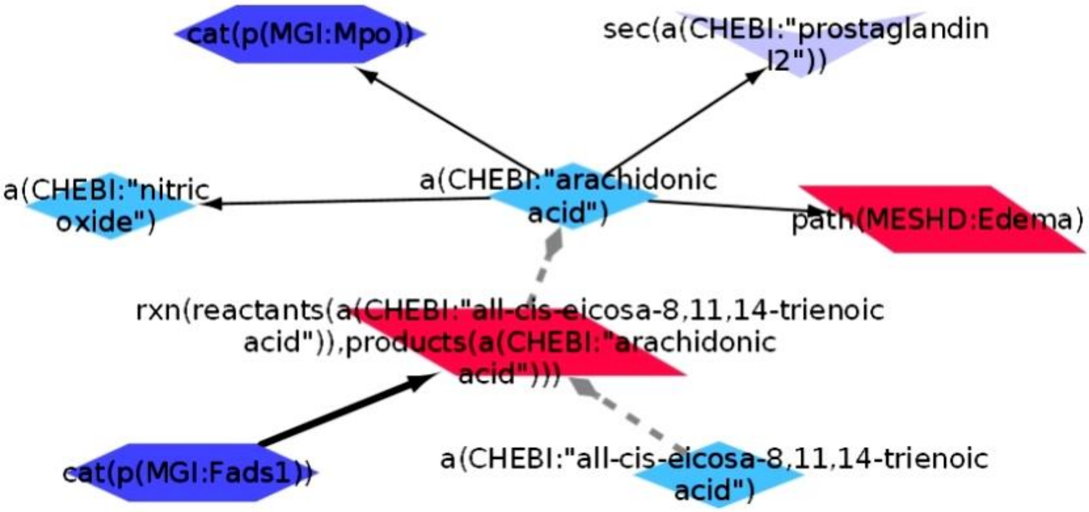

cat(p(MGI:Mpo))
sec(a(CHEBI:"prostaglandin I2"))
a(CHEBI:"nitric oxide")
a(CHEBI:"arachidonic acid")
path(MESHD:Edema)
rxn(reactants(a(CHEBI:"all-cis-eicosa-8,11,14-trienoic acid")),products(a(CHEBI:"arachidonic acid")))
cat(p(MGI:Fads1))
a(CHEBI:"all-cis-eicosa-8,11,14-trienoic acid")

Figure 3.5: Example of BEL Model with different reactions.

A qualitative modeling approach to model the disease is agent based modeling which has recently gained popularity and fulfills our requirements to model some aspects of MS.

3.11 Agent based modeling and its application in biomedicine:

Agent based modeling (ABM) is another approach to model complex systems**;** it is also known as individual based modeling. In this approach, each entity or agent is individually controlled by different parameters. It therefore has the advantage of providing attributes specific to a specific agent, depending upon the interactions of that agent with other agents as well as with the local environment. The power of individuality in this approach can be extensively used in systems biology, as in biology each cell or molecule has a specific role. Take, for example, antigens and antibodies; without describing their specific features as molecules, it would be much more complicated to simulate an environment in which they play an important role. Although an agent can represent a molecule, a cell or a complex, it is most commonly recognized and practiced that an agent represents a cell [183]. A typical ABM simulation would be cells (agents) interacting with each other in the local environment according to predefined rules, which are usually formed from knowledge gained via experiments and studies that is then translated into computational algorithms. Based on those rules, the output of the system will be justified; and if the relevant details are not sufficiently encoded, then the system may not provide the identical results. The results may vary from the encoded knowledge, thus including an iterative process to enrich the rules is a mandatory step for ABM. The data to feed those rules could come from in-vitro, in-vivo or any other experiments and studies; however, the data must be validated first. ABM has been used in the modeling of different diseases, one of which is cancer. ABM was also used to simulate both the growth of a brain tumor [184] and the role of heterogeneity in drug resistance [185].

3.12 Modeling neurological diseases and particularly MS:

As previously discussed, the brain is one of the most complex structures of the universe; thus, in order to model that level of complexity, a framework which could handle a similar level of complexity is needed. Systems biology provides that multi-level platform to integrate, analyze, and simulate models and datasets from different –omics studies. In recent times, neuroscientists have worked only with reductionist approaches by classifying the brain based on functionality, cellular composition, and parts. Even though a reductionist approach yielded some success, it is certainly not the way forward in finding a cure for multifactorial neurological diseases. For example, one cannot study memory, learning, and behavior by only observing neurons or any other individual cell type of the brain. Many fields such as neuroinformatics, computational neuroscience, and neurophysiology aim to decode brain functionality. These fields are shifting gradually towards systematic approaches, but they are not yet considered to be part of systems biology [186]. The ultimate objective of systems biology in neurodegenerative diseases is to find pathways involved in disease pathogenesis; this would be done by analyzing networks constructed based on different datasets e.g. gene expression, proteomic, and neurobiological experiments designed to aid in drug discovery. Many neurological disorders' models have been developed already, e.g. Alzheimer's disease model [187,188] and Parkinson's disease model [188].

3.13 MS disease modeling:

As discussed in the previous chapter, MS is a complex disease, with heterogeneity being one of its most complicated aspects. The heterogeneity of the disease has been modeled extensively, biologically (as shown in table 3.6) [189], but only with limited scope computationally (e.g. using stochastic Petri nets [190] and regression models [191]).

	Model system	**Strain: antigen**
Clinical course		
Relapsing–remitting	Wild type mice	SJL/J: PLP$_{131-151}$
	Transgenic TCR^{MOG}	SJL/J: MOG$_{92-106}$ (spontaneous)
	Wild type mice, adjuvant specific	C57BL/6:MOG$_{35-55}$ (adjuvant Quil A)
	Wild type mice, antigen/adjuvant dose specific	C57BL/6: MOG$_{35-55}$ (low dose)
Secondary progressive	Wild type mice	Biozzi ABH: spinal cord homogenate
	Wild type mice	NOD: MOG$_{35-55}$
Lesion localization		
Spinal cord	Most EAE models	
Opticospinal	Transgenic TCR^{MOG} x IgH^{MOG}	C57BL/6: MOG (spontaneous)
Brain	IFNγ or IFNγR deficiency	Multiple
(± spinal cord)	Wild type mice	CBA/J: PLP$_{190-209}$
	Wild type mice	C3H/Hej: PLP$_{190-209}$
	Wild type mice	C3HeB/Fej: MOG$_{97-114}$
	Transgenic TCR^{MOG}	SJL/J: MOG$_{92-106}$ (spontaneous)
	Wild type mice (CD8 T cell clones)	C3H/Fej: MBP$_{79-87}$
Pathological pattern		
Pattern I/II	Most CD4-mediated EAE models	
Pattern III/IV	Wild type mice (CD8 T cell clones)	C3H/Fej: MBP$_{79-87}$
	Cuprizone-induced demyelination	C57BL/6, Swiss Webster
	TMEV-induced demyelination	SJL/J

Table 3.6: Various Model systems used for MS, edited version of [189].

The MS models available so far do not illustrate a broad picture of the disease, while there are comprehensive models available already for other neurological diseases. Thus far, there is no MS computational model which collectively describes the disease's mechanism, the interacting molecules in certain phenotypes, and the pathways involved in the disease's processes. The following points are the main reasons to model MS:

- To more effectively utilize the large and growing body of literature on the disease, which is currently very difficult to follow or use to gain an overview of the disease's mechanisms
- To organize and clarify the heterogenic aspects of the disease, which would help answer questions pertaining to certain disease subtypes and mechanisms
- To identify and establish hubs of molecular interactions, key players in different disease processes, and other important pathways
- To assist in the discovery of potential biomarkers and drug targets

Various different approaches and methodologies have been used to model MS. Statistical modeling and Bayesian approaches have been proposed and applied in the classic manner to model the disease course and the heterogeneity of MS. Some of the examples of said approaches are nonlinear model of MS disease [192], and modeling time series of MS disease course [193]. Modeling has also been used in a longitudinal study of RRMS to illustrate the prevalent physical inactivity due to disease severity [194]. In another, similar longitudinal study, flexible modeling was used to measure the association of past relapses and disability occurrence in RRMS in order to help establish a prognosis of disability and disease progression [195]. Markov models have also been developed to link MS disease progression with age and to predict disability progression [196,197]. The Bayesian approach, using Expanded Disability Status Scale (EDSS), has been used to model disability progression and accumulation in MS [198]. A best-fitting model was also developed, longitudinally, for disease progression based on EDSS observations [199]. Binomial regression models and other statistical modeling approaches were used to compare the efficacy of different drugs (fingolimod, teriflunomide and dimethyl fumarate); it has been shown that fingolimod is more efficacious than the other two drugs [200].

Pennisi et al., [201] developed an agent-based model which demonstrates the oscillatory behavior of RRMS, using virtual patient data, and which illustrates the protective role of vitamin D in MS [202]. In addition, an extended model was developed to emphasize the potential role of blood brain barrier in treatment of the disease [203]. MRIs and image data were used to develop a model of brain atrophy which correlate deep grey matter atrophy with white matter abnormalities and cognitive functions impairment in RRMS [204]. A lesion formation model was also developed based on MRI data over the period of one year and analyze different patterns of T2 lesions [205]. Another model shows the pattern of depressive mood in RRMS over the period of 2.5 years, showing that certain factors such as marriage, older age, employment, and physical activity play a significant role in depression symptoms' development [206]. As far as the author's knowledge, no model of MS using systems biology approaches or molecular interaction maps of the disease is currently available; thus, our work is first of its kind and novel in its direction.

4 Chapter 4: Methodology

The methodology section has been split into the following two parts due to the different scopes of the work:

- Information retrieval and representation
- Modeling of MS disease

4.1 Information retrieval and representation:

The overall accomplishment of the work of information retrieval and representation is to build a functional MS ontology as per the standards of basic formal ontology (BFO) [207], which could support information retrieval and analysis, and integrate in SCAIView [28]. To accomplish this task, a methodology for semi-automated enrichment and translating ontologies was derived [29] and a program to develop ontologies with associated metadata [208] was developed. In addition to integration into a biomedical search engine, the ontology has also been used to retrieve clinical trial records for electronic health record (EHR) mining and to find co-stimulatory pathways in different neurodegenerative diseases (details in results section). Following are the steps which have been followed and the method used to achieve these results. All the relevant work has been published.

The first step in modeling a disease is to have access to relevant data associated with the disease across different resources e.g. clinical data, scientific literature and drug databases. A recent issue of Nucleic Acids Research reported a collection of 1,552 molecular biology databases [209]. In certain sub-domains of molecular biology, the data availability is not the issue but rather interpreting it; for example, modern -omics technologies produce a high volume of data with each experiment. Ontologies play a crucial role in integrating different databases and retrieving relevant knowledge. Disease ontologies have recently been used to represent domain specific knowledge [210,211]. These ontologies help to retrieve and gather all the relevant information about the disease from different platforms e.g. literature from PubMed, clinical information from EHR, and drug-specific knowledge from DrugBank and other domain-specific databases.

The objective of this section is to develop an MS disease ontology which could support the integration of disease-specific knowledge across different platforms and answer complex queries. The queries would be much more complicated than simply typing the disease term into

PubMed, as the ontology would facilitate precise information retrieval due to the associated semantics. In the following sections, I describe our methodology of developing an ontology and enriching it with one of the largest biomedical repositories available.

4.1.1 Foundational work for ontology development:

The first step to develop ontology was to collect all the concepts associated with the disease. For this two different approaches have been used. The first was manual collection of the disease-specific concepts by e.g. reading literature and websites in order to gather the concepts; for example, gathering all the relevant concepts mentioned on the website of Encyclopedia of Multiple Sclerosis [212]. In addition to looking at only one particular website, we were interested to do the concept collection task automatically and to generate a relevant text corpus from the internet so that corpus could be processed offline, and concepts could be tagged and extracted by using natural language processing (NLP) tools. With this purpose in mind, a methodology has been derived [30] by which one can rapidly generate a clean corpus from different websites. The description of the methodology is discussed later in this chapter.

The second approach was to use NLP tools (Named Entity Recognition) on some of the MS books, tag all the biological concepts, and then determine their role in the disease. We used the Temis Luxid tool (version 6) [6] and its biological entity recognition cartridge, developed by the Fraunhofer institute, to process the books. The tool chosen to develop the ontology was Protégé (version 3.3.1) [12], because of its usability and variety of available plug-ins. Following hardware specifications were used for all the work: Dell Latitude E4310 - Core i5 Notebook with Dual core processor, 4 gigabytes RAM, and 120 gigabytes hard drive.

After collecting the disease-specific concepts, we wanted to have all the synonyms associated with those concepts in order to broaden the coverage of the ontology. Synonym enrichment prevents information loss due to the various naming conventions in different literature sources; for example, Interleukin-17 could be written as IL-17 or IL17. The ontology must be rich enough to have as many synonyms of each molecule associated with the disease as possible, thus Unified Medical Language Systems (UMLS) was selected.

Ontology enrichment is a process of embedding metadata associated with the concepts defined in ontology. The specific metadata or attributes are added to have a unique set of concepts in an ontology specified for a domain. Different types of attributes are added to cover different aspects; definitions are added to have a common understanding of concepts; synonyms are added to have a broad coverage of ontology; and references are given to provide the source of knowledge. In addition, there could be as many attributes as developer wants e.g. date, comments, language, label, contributor, creator, identifier, etc. Manual ontology enrichment has several disadvantages which deter its usage being common. It requires a great deal of human effort and time, making it less attractive for scientists. Additionally, searching different sources for different concepts could lead to disagreements among concepts within an ontology and limit its application on a specific domain. There are few automated ontology enrichment tools, and these require technical expertise of computing and natural language processing; thus they are not a first choice for biologists. In addition, since most of the tools work on a corpus to generate a hierarchical ontology, the results could vary significantly based on the content of the corpus. These tools can help to develop de novo ontologies, but they may not be appropriate for enriching a domain-specific ontology. To have a harmonized ontology it is a good practice to retrieve attributes of concepts from a unified and broadly accepted database. Using automated tools to query the repository could reduce human errors and make this repetitive and time-consuming task easier. Further, it is also a good practice to assign unique identifiers to each concept in order to make them interoperable with and semantically relevant to other ontologies.

UMLS is a very large repository of medical concepts which integrates and streamlines many health vocabularies to enable interoperability among them. UMLS has more than 100 source vocabularies [213] and it has been reported that the 2009AB release of the UMLS Metathesaurus contained 2,120,271 biomedical concepts and 5,305,932 unique terms [16]. Two areas of its usage are in electronic health records software development and by health-related language translators. UMLS deals with the complexity of different biomedical concepts by assigning a unique identifier to them, called a Concept Unique Identifier (CUI). CUIs have a unique alpha-numeric (C0000000) format which is consistent, though the sources of the concepts may differ. CUIs are being used to map concepts from different sources to UMLS. Mapping ontology concepts to UMLS CUIs makes them more interoperable, accessible and provides a common understanding of the concepts. By using UMLS CUIs, additional metadata of the concepts (definitions, synonyms, etc.) could also be integrated.

4.1.1.1 Setting up UMLS locally:

UMLS can be queried via the web; however, in order to query hundreds of concepts it was necessary to set up a local instance of UMLS. It is possible to load UMLS onto a local database via various configuration scripts i.e. MySQL, Oracle, or Microsoft Access [214], and thus expedite data querying and retrieval. MySQL Server 5.5 has been used because of the stability and free availability. The installation and configuration were done following the guidelines given at [215]. The UMLS database (when implemented on MySQL) is approximately 26 gigabytes, which is fairly large for the average personal computer. This contrasts with the database size restrictions in some programs such as Microsoft Access, which can only create a database with the maximum size of 2 gigabytes [216]. Due to the large size of UMLS, the performance of the system running locally is often compromised and manual querying is not considered a preferred choice. One solution for this problem is to set up an automated loop to query, retrieve, and store data, thereby repeatedly and regularly querying hundreds of concepts in a short time. This can be done via automated workflow programs e.g. Taverna [217] and Konstanz Information Miner (KNIME) [5] and discussed in details in following sections.

4.1.1.2 Connecting KNIME to MySQL:

KNIME is an open source, easy to use and graphical user interface workbench for different data analytics processes. It provides a broad range of nodes and plug-ins to connect to web services, run scripts and execute external applications within the workbench [5]. We wanted to query UMLS implemented on MySQL database, and with KNIME (version 2.8) it was relatively easy to connect to different databases including Oracle, SQLite or any other JDBC/ODBC-compliant databases by using “Database reader node” (Table 4.1). The connector “mysql-connector-java-5.1.12-bin.jar” was downloaded from MySQL website [218].

Configuration of KNIME “Database reader node”	
Database Driver:	Com.mysql.jdbc.Driver
Database URL:	Jdbc:mysql://localhost:3306/umls
User Name:	Root
Password:	Root

Table 4.1: MySQL connection settings in KNIME “Database reader node”.

4.1.1.3 Automated data retrieval from UMLS:

The list of disease-specific concepts gathered from the processed corpus of online MS Encyclopedia and MS books was used to query the UMLS via an automated workflow of KNIME (Figure 4.1). The starting node "XLS reader" reads the list of concepts and formulates it as KNIME standard table output. The second node "TableRow to Variable Loop Start" uses each row of a table to define a variable for loop iteration. The third node "Database Reader" connects to the database (UMLS) and queries with the variable provided in the previous step (Node 2). The fourth node "Loop End" continues the loop until the last row of the file and ends the loop. The last node "CSV writer" writes the output as comma-separated values (CSV) file. This automated approach queried each concept one by one and gathered the associated data into an output file.

Most of the retrieved results were accurate because quotation marks were used for more than one word concepts and acronyms were avoided. After the data retrieval, we wanted to integrate the automatically retrieved results into our ontology, and this was done by using OntoFast (version 01) [208].

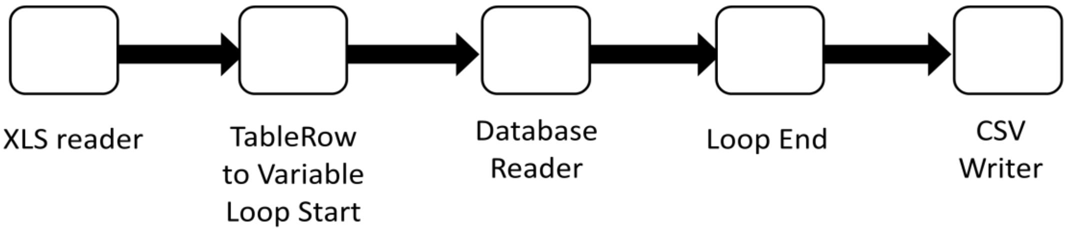


Figure 4.1: KNIME workflow to automate the querying process from MySQL (Loaded with UMLS). (From left) The first node reads the Excel table, second node creates a loop, third node reads the each item of the loop and queries the databse, fourth node repeats the loop until it ends, and fifth node writes all the data to CSV file.

4.1.2 Automated ontology translation from UMLS:

The limiting factor of ontologies' usage is their availability in only few languages. Most of the ontologies are available only in English, which restricts their application on English datasets only. The biomedical domain is one of the largest domains which has many well-designed ontologies as well as a widespread application. Bioportal [219], a biomedical ontology database, has 370 ontologies at the time of this writing. As discussed above, most of them are only available in English, thus it is practically impossible to use them in conjunction with other language datasets to retrieve information from other language repositories. Here we describe our use case of ontology enrichment with the Spanish translations of concepts by using UMLS in a semi-automated way. The aim of our work was to extract knowledge from the EHR dataset, which was only available in the Spanish language.

There are some translation tools already available which could translate an ontology into another language. One such tool is LabelTranslator [220], which translates via Google translate and other web services, and requires manual selection of the individual, respective translated concepts. On the contrary, our system offers a much easier integration of translated concepts queried from an authentic domain source. Furthermore, our approach is very easy to use and requires only copy-and-paste to expand the ontology with concepts available in other languages. Figure 4.2 shows the overall outline of our approach for translating concepts and enriching ontology with them.

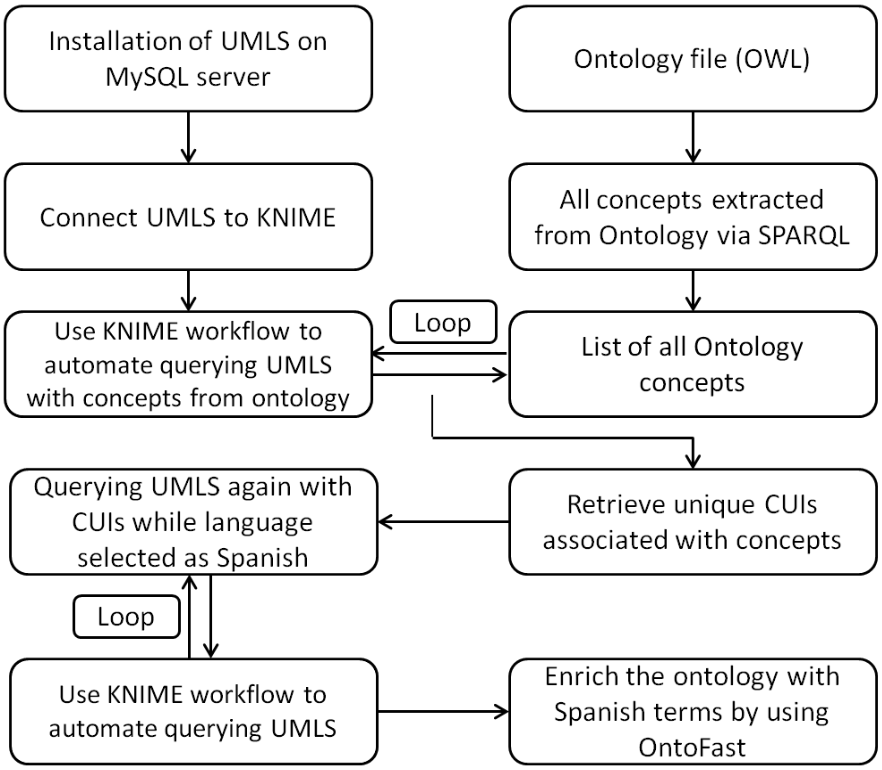

Installation of UMLS on MySQL server
Ontology file (OWL)
Connect UMLS to KNIME
All concepts extracted from Ontology via SPARQL
Use KNIME workflow to automate querying UMLS with concepts from ontology
Loop
List of all Ontology concepts
Querying UMLS again with CUIs while language selected as Spanish
Retrieve unique CUIs associated with concepts
Loop
Use KNIME workflow to automate querying UMLS
Enrich the ontology with Spanish terms by using OntoFast

Figure 4.2: Overall outline of the Ontology enrichment and translation. Left side workflow starts with the configuration of the UMLS system while right side starts from Ontology file. After extracting all the ontology concepts, they can be queried over UMLS to retrieve associated metadata. Retrieved concepts can be simply embedded into ontology for enrichment or used further to get relevant concepts in other languages.

In addition to the automated enrichment and translation methodology, a program has been developed to quickly develop ontologies with all their associated metadata. Details of the program will be discussed in results section of the thesis.

As discussed above, In addition to text mining MS books and scientific papers, another approach to gather concepts is to generate datasets from MS-relevant websites by mining the web and then processing those datasets with NLP tools. A methodology has been derived to obtain clean datasets from the web, which were scattered across different web pages, in a short amount of time.

4.1.3 Corpus generation from the web:

This method requires a plug-in oriented web browser e.g. Mozilla Firefox [221] or Google Chrome [222] as well as the DownThemAll plug-in [223], the Link Gopher plug-in [224] or any variation of GREP Program [225]. DownThemALL is a browser extension which can download hundreds of documents in one go. GREP is a widely used utility program which provides different functionalities, but for the task it was used only to filter results.

All the programs need to be installed on a PC, and then clean datasets from different websites can be downloaded in a relatively short period of time and with little effort. Please note that some websites may prevent massive downloading and are only available for human access, thus prior permission may be required. This technique has been used on LinkedCT.org, one of the largest semantic repositories of clinical trials, which processes ClinicalTrials.gov and transforms it as RDF/XML. For this work, following versions of software were used:

- Mozilla Firefox version 17
- DownThemAll version 2.0.16.1-signed
- Link Gopher version 1.3.2.1-signed.1-signed

The initial steps of the corpora creation require identification of the pattern of the hyperlinks of the data you are interested in. If the links are available on one page, then DownThemAll can automatically detect them and you can start downloading the dataset or web pages instantly. If the actual data is beneath a few layers of web pages, then you can download first the source page(s) and then the actual data itself. This is done by combining all of the source html pages and extracting hyperlinks via Link Gopher or by using the GREP program. The good feature of

GREP is that it will also bring the data within the proximity of up to 5 lines from the actual search term, which could help with understanding the pattern of hyperlinks.

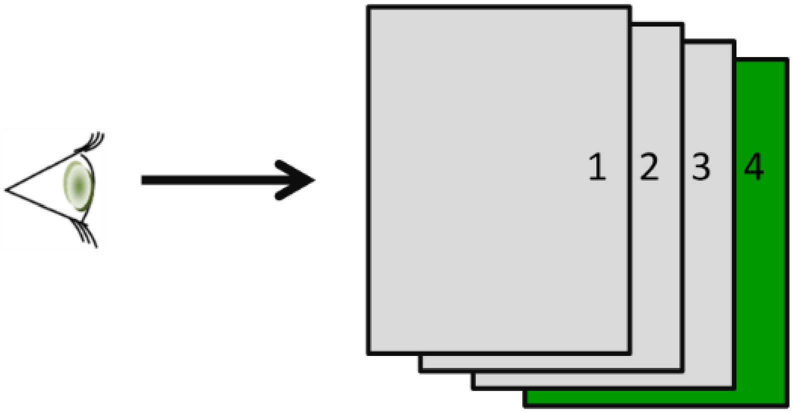


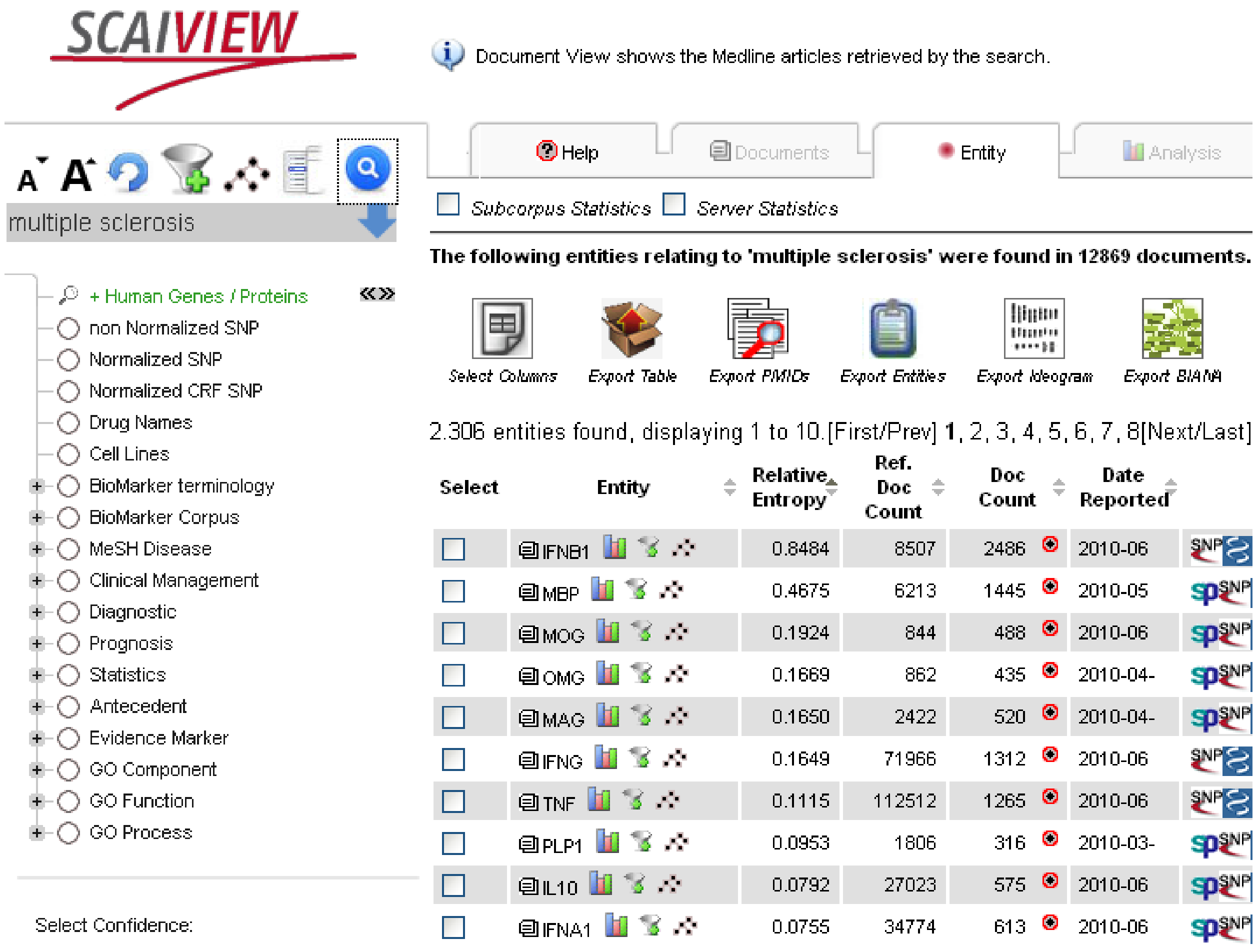
Figure 4.3: A perspective from user interface. The actual dataset was concealed below 3 web pages (shown in green), after search results are displayed on the website.

In the presented scenario, the dataset was beneath many other pages, and the hyperlinks of the actual files were scattered across different pages. As mentioned above, DownThemAll was used to collect the top pages (all three layers) and different terms associated with the same disease (Figure 4.3). The first layer contained the name of the disease with different orders and synonyms. There were 11 pages on the second layer which referred to 175 pages (third layer) of relevant data files, but in html format. The third layer also contained the link to the actual data file (in RDF format) as a hyperlink. All 175 pages were collected and patterns of hyperlinks, pointing to the actual data files, were observed. The manual work would have taken too long for this task, as one would have had to click forward and backward hundreds of times repetitively for quite some time. With the help of tools mentioned above, the task was done relatively quickly. This task helped us to collect MS relevant concepts from web, which were not found in MS books e.g. Turmeric.

After ontology development and enrichment, the next step was to develop the disease model and then make this model available within the scientific community, thus enabling iterative updates. It is worth mentioning that the MS ontology has been integrated into SCAIView, a semantic biomedical search engine. The concepts which were used in ontology development were also used to build disease models.

4.2 Modeling of MS disease:

The second part of our methodology discusses the approaches of modeling that we used. We used a bottom-up systems biology approach to develop a molecular interaction map of the disease. We used a semantic biomedical search engine (SCAIView) to rank molecular entities associated with the disease. The aim of the model is to have a graphic representation of the disease mechanism, and by looking at it one can immediately get to know important molecules, drugs acting on certain pathways, and phenotypes caused by different interactions. This map must provide an established knowledge of the disease, so that one does not need to sift through thousands of published papers or a large body of literature. The semantic tool, SCAIView, processes all the PubMed citations and provides filtration of results based on certain terms e.g. human genes and proteins etc. (Figure 4.4). This makes it convenient to distinguish knowledge between humans or any other organisms.



Document View shows the Medline articles retrieved by the search.

multiple sclerosis

□ Subcorpus Statistics □ Server Statistics

The following entities relating to 'multiple sclerosis' were found in 12869 documents.

2,306 entities found, displaying 1 to 10.[First/Prev] 1, 2, 3, 4, 5, 6, 7, 8[Next/Last]

Select	Entity	Relative Entropy	Ref. Doc Count	Doc Count	Date Reported
□	IFNB1	0.8484	8507	2486	2010-06
□	MBP	0.4675	6213	1445	2010-05
□	MOG	0.1924	844	488	2010-06
□	OMG	0.1669	862	435	2010-04-
□	MAG	0.1650	2422	520	2010-04-
□	IFNG	0.1649	71966	1312	2010-06
□	TNF	0.1115	112512	1265	2010-06
□	PLP1	0.0953	1806	316	2010-03-
□	IL10	0.0792	27023	575	2010-06
□	IFNA1	0.0755	34774	613	2010-06

Figure 4.4: SCAIView user interface and top ranked results based on Relative Entropy are shown. Top left shows the search term "multiple sclerosis" and filter "Human Genes / Proteins".

The platform chosen to model this knowledge is a Systems Biology Workbench tool called CellDesigner (version 4.1) [1]. Other tools were also evaluated, including Cytoscape[226], but CellDesigner was preferred due to its various features. CellDesigner is a state-of-the-art structured diagram editor for drawing gene-regulatory and biochemical networks. Its intuitive user interface helps draw diagrams in rich graphical representation with personalized design. Networks are constructed based on a state transition diagram proposed by Kitano et al. [2]. Recent versions further comply with SBGN process description diagrams [227]. Designed as a standalone tool, this powerful software is network-aware and therefore can connect to several major databases (DBGET[228], SGD[229], iHOP[230], Genome Network Platform[231], PubMed[232], Entrez Gene[233], SABIO-RK[234]) as well as retrieve models from BioModels.net[235]. CellDesigner lacks the wiki integration of PathVisio[236], has limited network analysis capabilities, and has fewer available plug-ins than Cytoscape. Furthermore, its source code is not available. Despite these shortcomings, the appealing user interface, the native support for SBML and the straight integration with the System Biology Workbench [237] were sufficient reasons for making CellDesigner our final choice as the tool to be used for our work. CellDesigner uses MIRIAM [8] for annotation of the SBML models. The version of the MIRIAM database has been extended with additional data types for storing sentences, for annotating with UMLS and with MEDDRA [238].

4.2.1 Molecular interaction map of MS:

After developing the ontology, some of the concepts which were used in ontology have been used further to develop molecular interaction map of MS. As discussed above, the objective of the work was to have a disease map with sufficient details of the disease mechanisms, drugs targeting different pathways, and the clinical outcome of the interaction between different molecular entities. One specific requirement was to have a supporting sentence behind each edge of interaction with the PubMed-ID (PMID) or source of that sentence.

4.2.1.1 Selection of seed entities:

The topmost occurring entities associated with the search term “Multiple sclerosis” and “Human genes / proteins” (Figure 4.4) in the SCAIView have been selected. The selection of proteins and genes was made based on the Relative Entropy score (a confidence measure) above 0.050. Then all the abstracts associated with the molecules and MS disease association with any of the molecules were read to find an interaction. The molecules which were not in the ranking, but were mentioned in the abstract as interacting with one of the ranked molecules were also extracted and used in interaction map.

4.2.1.2 Selection of corpus:

A corpus of text has been collected after reading the abstracts. The corpus (collection of abstracts from which the sentences and relations were taken) was collected by considering only those abstracts which contained “Multiple Sclerosis” or its animal model “Experimental Autoimmune Encephalomyelitis” and had co-occurrence with entities ranked by higher relative entropy in SCAIView. The corpus must also contain either information on the molecules involved in certain disease states, or on their interaction manner in relation to disease behavior changes (either increasing the likelihood of disease pathogenesis or suppressing the phenotypes associated with a disease state).

4.2.1.3 Corpus annotation:

Sentences from the corpus were manually tagged and extracted where a clear statement of relation between any two entities (gene, protein, drug, simple molecule or phenotype) was mentioned. The entities were standardized by being given a UMLS CUI, in addition to the name mentioned in literature. The CUIs were also stored under the MIRIAM section of each node of the map.

4.2.1.4 Disease model development:

The disease model was developed by linking the nodes (molecular entities, drugs, phenotypes etc) with edges supported by sentences, which were embedded within the model under each edge. This embedding also contained the source information or PMID to track back to the original abstract. The local version of CellDesigner (4.1) was extended to accumulate text annotation. To re-use this embedded knowledge and make it available for the scientific community, a plug-in was developed to transform the disease model into a Mediawiki knowledge base [13]. The model can be curated by groups using Payao web portal [239].

4.2.1.5 Validation of model:

Validation of model was done by reading reviews which were published recently about the MS disease and those were:

(1) Therapy of MS [240]

(2) Multiple sclerosis therapies: molecular mechanisms and future [241]

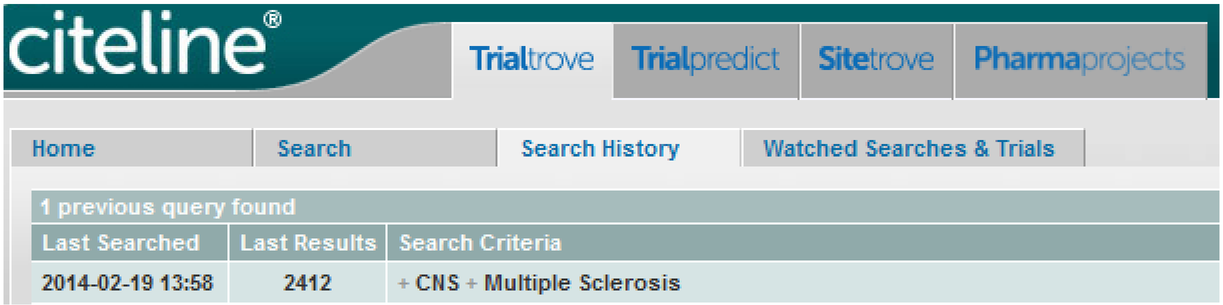
(3) Multiple Sclerosis: risk factors, prodromes, and potential causal pathways [242].

The validation was needed as one could argue that most of the co-occurring molecules associated with MS in SCAIView or in scientific literature are generally the longer known molecules. By using that approach one could easily miss the novel and recent discoveries. The recent discoveries may not be included in as many scientific publications, as compared to a molecule discovered 10-20 years ago. So, recently published reviews were found which covered different aspects of MS, and compared their findings with our model. Newly found entities in reviews were added by searching entities (e.g. xxx) with “Multiple sclerosis” i.e. “Multiple sclerosis + xxx” in SCAIView and the retrieved data was added to the disease model.

The clinical aspect of the disease and drugs data were missing from the model as so far the knowledge was only extracted from established scientific publications thus next clinical data and drugs mode of action were overlaid on the model.

4.2.2 Overlaying clinical trials and DrugBank data:

The data from public and commercial clinical trial datasets were integrated into the model. A commercial dataset was taken from Trialtrove, which is one of the largest repositories of clinical trials [243]. 2,412 clinical trials were found in Trialtrove by using the search term “Multiple Sclerosis” under the category of therapeutic area “CNS” (Figure 4.5).

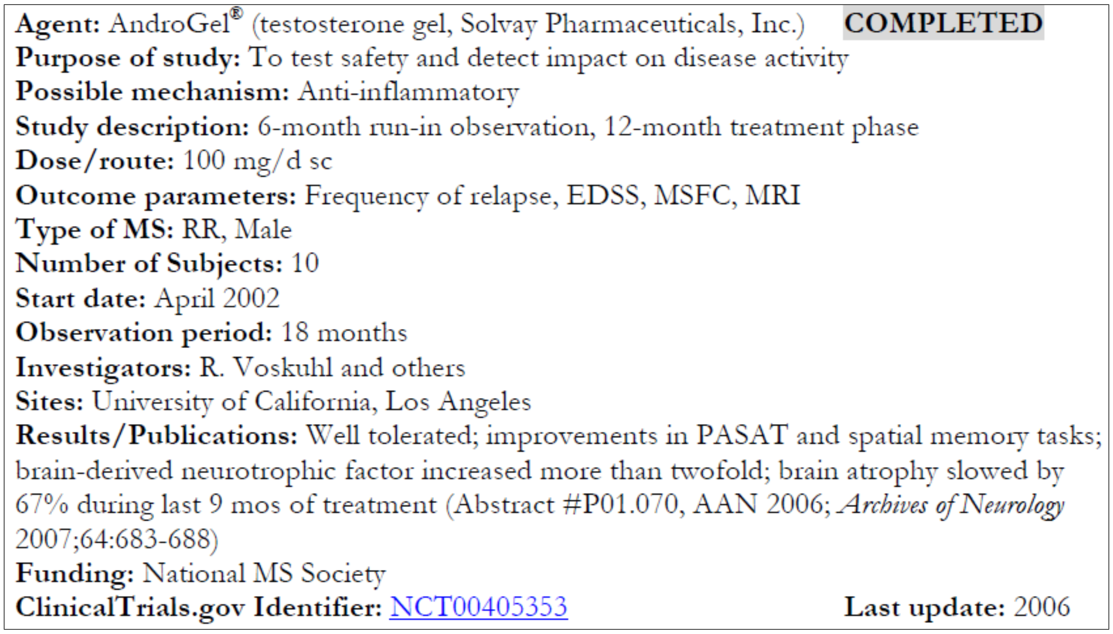


1 previous query found

Last Searched	Last Results	Search Criteria
2014-02-19 13:58	2412	+ CNS + Multiple Sclerosis

Figure 4.5: Clinical trials associated with Multiple Sclerosis in a commercial database [243].

Then molecules associated with RRMS, a subtype of MS, were filtered out and their associations were studied. In addition to Trialtrove clinical trial data, the public clinical trials dataset from the National Multiple Sclerosis Society’s (NMSS) [244] was also obtained and processed. The NMSS clinical trial dataset is unique in a way that it is the only public dataset (to the author’s knowledge) which contains the results section and name of the publications (Figure 4.6).



Agent: AndroGel$^{®}$ (testosterone gel, Solvay Pharmaceuticals, Inc.) **COMPLETED**
Purpose of study: To test safety and detect impact on disease activity
Possible mechanism: Anti-inflammatory
Study description: 6-month run-in observation, 12-month treatment phase
Dose/route: 100 mg/d sc
Outcome parameters: Frequency of relapse, EDSS, MSFC, MRI
Type of MS: RR, Male
Number of Subjects: 10
Start date: April 2002
Observation period: 18 months
Investigators: R. Voskuhl and others
Sites: University of California, Los Angeles
Results/Publications: Well tolerated; improvements in PASAT and spatial memory tasks; brain-derived neurotrophic factor increased more than twofold; brain atrophy slowed by 67% during last 9 mos of treatment (Abstract #P01.070, AAN 2006; *Archives of Neurology* 2007;64:683-688)
Funding: National MS Society
ClinicalTrials.gov Identifier: NCT00405353 **Last update:** 2006

Figure 4.6: Snippet of a clinical trial summary record available on the NMSS dataset [244].

DrugBank RDF dataset [245] was used to extract MS drugs' mode of action and the molecules which the drugs act upon. This knowledge was also mapped into the disease map.

One challenging aspect of MS is the different disease progression patterns of its subtypes. To reveal the patterns behind that aspect, the biomarkers specific to each subtypes were extracted from literature so the specific molecules for each subtype can be revealed.

4.2.3 Time series of MS disease progression:

The complexity behind MS disease subtypes and their progression may be discovered by identifying biomarkers specific to each disease subtype, segregate different disease subtypes, and isolate and mark the clinical endpoints of a particular disease subtype in chronological order, with the help of associated biomarkers. A rigorous search was performed to select a specific body of literature, including full text research papers. An advanced text mining tool Temis Luxid (version 6) [246] was used, as well as text mentioning the association of any molecules and their significance to the disease subtype was manually extracted. After careful selection of contents, 560 research papers have been selected.

4.2.3.1 Disease segregation based on biomarkers and therapeutic agents:

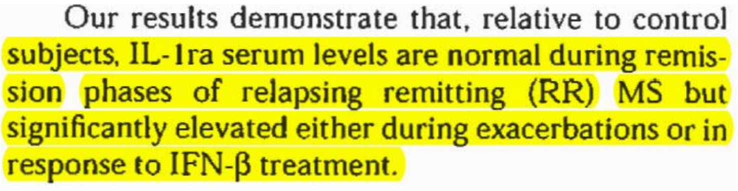
The results and observations of patients in different disease stages i.e. Clinical Isolated Syndrome - CIS (even though it is not considered as disease subtype, it may be considered as prediction stage of the disease), RRMS, SPMS, PPMS and PRMS have been segregated.

One of the challenges in the work was to find a time-dependent shifting of molecular behavior in relation to disease phenotype as well as to find out potential perturbed pathways involved in each of the stages. The question has been asked if it is possible to discover a means to control or prolong the shifting phase and reduce disease severity by optimizing drug combination. This could help us with finding events which cause the disease worsening in different time intervals between the different stages. Furthermore, we were also interested to identify biomarkers which play a role in acute types and rapid progression of the disease. In addition to the four subtypes of MS, we also looked at the molecular mechanism of CIS. CIS is the primary indication before the disease starts getting worse, and this is considered one of the basic indicators of the start of disease (MS). As mentioned above, it is not a disease sub-type as the occurrence of MS after CIS is 50% [86]; however, looking at the biomarkers associated with CIS could help predict the possibility of MS occurrence.

The four subtypes of MS are known, but the longitudinal sequence of the events with respect to disease progression is not necessarily linear. Sometimes the course of the disease is aggressive and it rapidly changes from one point to another. Some patients experience no PPMS stage, but rather progress directly from RRMS to SPMS. As for covering the widely known knowledge and patient population, a pattern of linear progression has been considered.

4.2.3.2 Annotation of the scientific literature:

Important scientific findings in full research papers (PDF format) were annotated, and highlighted text was extracted with Zotero (version 3.0.9) [247]. Interpreted results were collected in a Microsoft Excel table and scanned research papers were processed before annotation with an optical character recognition (OCR) software. Figure 4.7 shows the text snippet of an annotated research paper.



Our results demonstrate that, relative to control subjects, IL-1ra serum levels are normal during remission phases of relapsing remitting (RR) MS but significantly elevated either during exacerbations or in response to IFN-β treatment.

Figure 4.7: Example of the annotated Text of a research publication. Different parts of the publications were annotated; here the result section is shown.

The output table contains many columns with relevant findings and biomarkers found in literature. Data of different biomarker types (discussed in chapter 2) from 560 research papers with all MS disease subtypes were extracted. The resultant lists of biomarkers were used to develop biomarker molecular interaction maps. The enrichment of those maps was done using the Ingenuity Pathway database [4]. Enrichment brought the resulting pathway where those molecules play crucial roles. Ranking of pathways was done based on the number of molecules involved in each pathway.

In addition to biomarkers-based segregation, we also separated different disease subtypes based on the therapeutic intervention. There are different sets of therapies given to patients at different stages of the disease. These are known as different lines of therapies e.g. first line therapy, second line therapy etc. The drug regimens were selected based on the severity of the disease, and we hypothesized that since there are different pathways involved in each disease subtype, the targets being hit by different lines of therapies could also be different. If there is a coherence of those pathways, then the perturbed pathways certainly play a role in the disease; this would be an affirmation of our findings. To do this we used all FDA approved drugs available for MS and found the pathways they are targeting. This was done by constructing each drug network by populating the knowledge from Ingenuity pathway database and then mapping those networks to pathways.

After exploring various streams of knowledge of MS disease and laying the foundation of a system based on which a knowledge base can be established and simulation can be executed to generate hypothesis which may open the frontiers of novel drug targets, the new challenge was to have a qualitative model of the disease. The rationale behind developing an agent based model is that there is a large body of knowledge available which supports the notion that MS is disease of cellular interplay and shift from one cellular population to another may change the pattern of the disease. In the introduction part of the thesis, the protective factors and risk factors have been discussed and shown (Figure 2.2). In the next section, we will explore one aspect of cellular interplay and their role in disease.

4.2.4 Agent-based model of Treg-Teff interplay and its role in RRMS:

Agent-based modeling (ABM) meets the specification required by some of the biological mechanisms. The dynamics of certain biological entities can be embedded in ABM, such as position, function of time, internal states (e.g. age, active etc) as well as certain interactions such as binding, which modifies the behavior of the interacting agents. These dynamics of the global system are generated by interactions of all agents in a certain environment. ABM deals with heterogeneity and spatial issues of agents, and it is relatively easy to describe the complex rules assigned to the agents. Agent-based models have been used for simulating many diseases e.g. HIV [248–250], mammary carcinoma and lung metastases [251,252], atherosclerosis [253] as well as the cell-based immune response to cancer cell antigen presentation [254]. Since there were many disease-specific agent-based models, ABM has been selected to model the interplay between T-regulatory (Treg) and T-effector (Teff) cells and, additionally, their role in RRMS and causing relapses. To develop a model of RRMS, following assumptions were made based on the available experimental findings:

(A) The interplay amongst Treg-Teff cells and the up-regulation of Treg cells by Teff cell signaling [255]. The imbalance between the two cell types has been shown to play a crucial role in both MS and Type 1 diabetes [256,257].

(B) The inhibition of Teff cells by Treg cells by means of cell to cell contact inhibition [258] and immunosuppressive cytokine secretion [259].

(C) The inflammation caused by EBV, as it has a role in modulating the human immune system and is considered to be an important factor in MS pathogenesis [84].

(D) The role of biomimicry; EBV-specific T cells cross-react with auto-antigens such as Myelin Basic Protein (MBP) and destroy it.

(E) The correlation between relapse and neural damage, as some studies have shown the presence of biomarkers specific to axonal damage and myelin damage (NFL and MBP) to be higher in RRMS patients [260].

We used NetLogo programming language [261] and software suite (version 5.0.1), which implements an agent-based oriented programming language. It is an excellent option to model and simulate multi-agent environments and complex systems.

4.2.4.1 The model and simulation:

The model was developed in NetLogo using the following agents (called turtles in NetLogo):

- Epstein Barr Virus (EBV)
- Treg Cells
- Teff Cells

The model's environment was represented as a myelin sheath. Myelin is considered to be the place where all the interactions of cells and viruses take place. In the interest of simplicity, the blood-brain barrier was not shown. The following are the major events of the simulation:

4.2.4.2 Role of EBV:

EBV virus causes infection and invokes mimicry which activates the auto-reactive Teff cells and Treg cells [262]. Due to the antigenic mimicry of EBV epitope, Teff cells attack the myelin as if it were EBV. The viruses themselves do not interact with the myelin patches. A virus has a radius (virus_radius), and within this radius T cells can become activated; the virus will then be eliminated.

4.2.4.3 Role of Treg cells:

There are two states of Treg cells in the model, resting and activated. Resting cells do not interact with Teff cells; but rather, they become activated after EBV infection. Activated Treg cells can suppress Teff cells and duplicate themselves in response to positive feedback. The suppression of Teff cells occurs via different cytokines' signals released by Treg cells. In the interest of simplicity, the various cytokines were not modeled. The duplicated Treg cells have a life reduced by half, and the new cells are active and present in the vicinity of the same myelin patch.

4.2.4.4 Role of Teff cells:

As with Treg cells, there are two states of Teff cells in the model, resting and activated. Resting Teff cells do not interact with myelin or Treg cells. They become activated once they interact with the EBV. Activated Teff cells can damage myelin, due to mimicry, and duplicate. If activated Teff cells are in a patch with the myelin quantity higher than 0, then Teff cells attack the myelin and reduce the amount of myelin and duplicate themselves. If the myelin quantity is 0, then Teff will not be able to duplicate. The duplicated Teff cells will have their life reduced by half, and the new cells will be active and present in the same patch's vicinity.

Parameter	**Meaning**
treg_radius	max visibility radius of Treg
eff_dup	max. duplication rate of Teff
init_mye	initial quantity of myelin per patch
eat_mye	quantity of myelin destroyed by Teff
pt	max. duplication rate of Treg
patch_density	max. no of entities per patch allowed to have duplication
Teff_life	Teff mean half-life
Treg_life	Treg mean half-life

Table 4.2: Parameter used for the model and their meaning.

4.2.4.5 Role of environment or myelin:

The environment (the patches) represents a small portion of white matter and is initially grey in color. The variable used for this parameter is init_mye. The damaged caused by Teff cells to the myelin is categorized as recoverable or non-recoverable. The recoverable patch is also initially grey in color, while non-recoverable damage is black in color. The variable to define the damage is called ate_mye. The damage is either recoverable or unrecoverable, depending upon the availability of myelin in the vicinity as per rec_mye rate at every time-stamp; otherwise, the damage is unrecoverable. The recovery of myelin is based on the repair mechanism of Oligodendrocytes [263]. The time-stamp was given as 2.4 hours, as it allows a good degree of granularity to simulate single relapse and also allows reasonable disease progression in a simulated time span of five years (18,250 times). A random number is generated in the start of simulation and the value is then set to number of agents to have a randomized simulation. All agents are free to move and interact with each other.

5 Chapter 5: Results

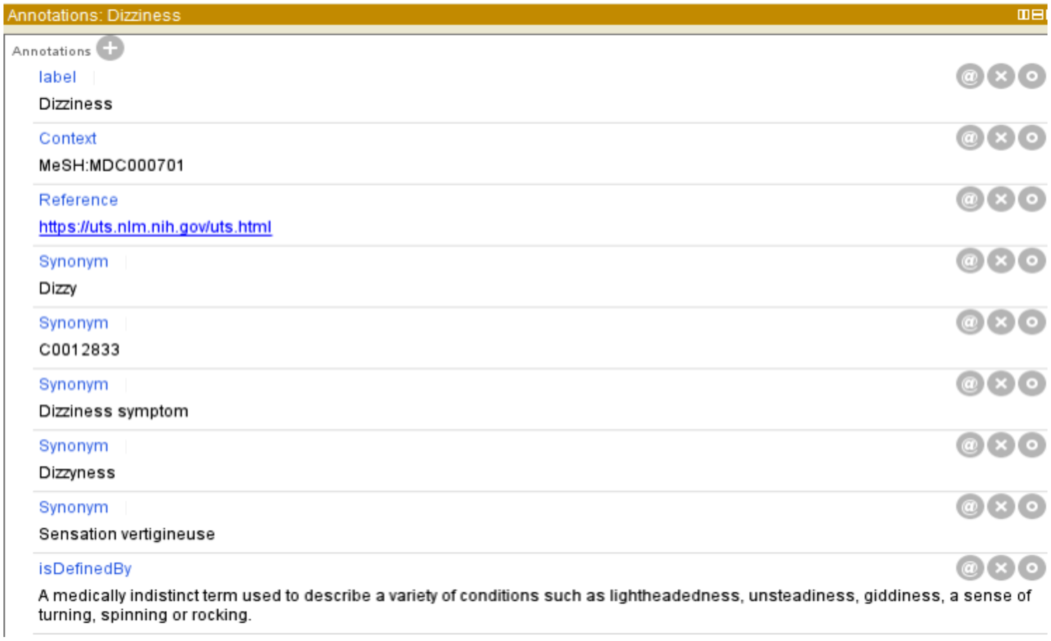
In agreement with the methodology section, the results section is also divided into the following two main parts:

- Information retrieval and representation
- Modeling of MS disease

5.1 Information retrieval and representation:

5.1.1 Ontology Development:

The purpose of MS ontology development was to have a unique set of terms associated with the disease to help integrate knowledge scattered on different platforms, and to create a common and unified understanding of the disease-relevant concepts. The scope of the ontology was to use it for disease specific information extraction and to model MS disease. The MS ontology is developed using standard basic formal ontology framework. The main classes of ontology are: 1) Clinical Presentations 2) Risk Factors 3) Molecular Entities. In addition to those classes there are also concepts about demographics and the social impact of the disease. There are 1,170 concepts in the ontology with 7,205 synonyms, which equals roughly six synonyms for each concept (Figure 5.1). Most of the concepts are molecular entities because of the application and scope of the ontology. A list of MS biomarkers used in ontology are taken from one of the recently published papers [106]. The resulting extraction associated with these concepts would be used to model the disease mechanism and the role of molecules in it.



Annotations: Dizziness

Annotations

label
Dizziness

Context
MeSH:MDC000701

Reference
https://uts.nlm.nih.gov/uts.html

Synonym
Dizzy

Synonym
C0012833

Synonym
Dizziness symptom

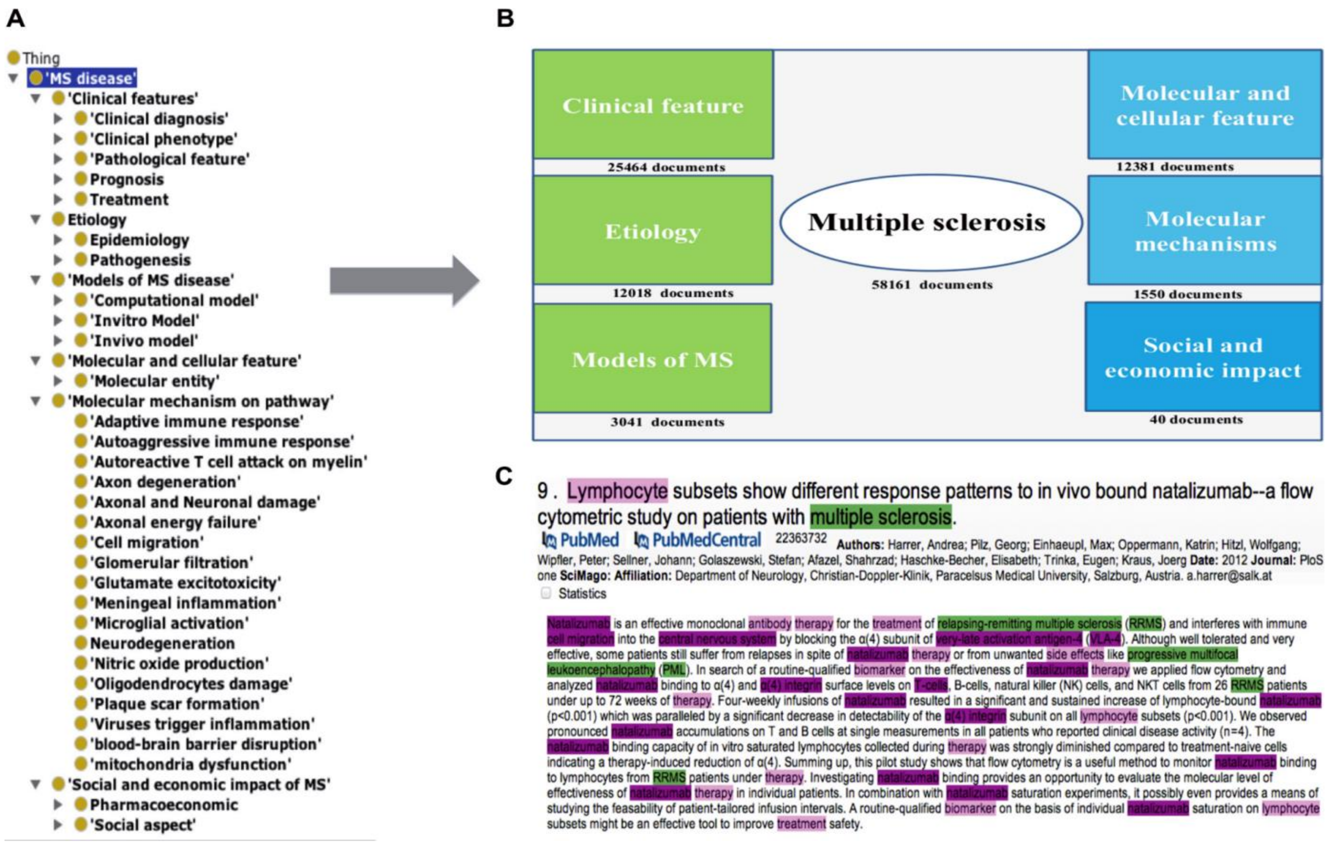
Synonym
Dizzyness

Synonym
Sensation vertigineuse

isDefinedBy
A medically indistinct term used to describe a variety of conditions such as lightheadedness, unsteadiness, giddiness, a sense of turning, spinning or rocking.

Figure 5.1: Each concept in ontology contains different types of associated metadata e.g. Definition, Context, Synonyms, and Reference.

Due to the large number of synonyms and well defined concepts with standardized identifiers, MS ontology (MSO) enables the retrieval of better results. The identifier makes it possible to reuse the same ontology across different websites and domains (the identifier is a unique alpha-numeric code from UMLS) and retrieve specific information from them. MSO also facilitates the discovery of co-morbidities associated with the disease, as demonstrated in the research paper [28]. Mining PubMed is another application scenario of the ontology. Another application is to use the concepts with all the relevant tagged knowledge to develop the molecular interaction map of MS.



9 . Lymphocyte subsets show different response patterns to in vivo bound natalizumab--a flow cytometric study on patients with multiple sclerosis.

22363732 **Authors:** Harrer, Andrea; Pilz, Georg; Einhaeupl, Max; Oppermann, Katrin; Hitzl, Wolfgang; Wipfler, Peter; Sellner, Johann; Golaszewski, Stefan; Afazel, Shahrzad; Haschke-Becher, Elisabeth; Trinka, Eugen; Kraus, Joerg **Date:** 2012 **Journal:** PloS one **SciMago:** **Affiliation:** Department of Neurology, Christian-Doppler-Klinik, Paracelsus Medical University, Salzburg, Austria. a.harrer@salk.at

Natalizumab is an effective monoclonal antibody therapy for the treatment of relapsing-remitting multiple sclerosis (RRMS) and interferes with immune cell migration into the central nervous system by blocking the α(4) subunit of very-late activation antigen-4 (VLA-4). Although well tolerated and very effective, some patients still suffer from relapses in spite of natalizumab therapy or from unwanted side effects like progressive multifocal leukoencephalopathy (PML). In search of a routine-qualified biomarker on the effectiveness of natalizumab therapy we applied flow cytometry and analyzed natalizumab binding to α(4) and α(4) integrin surface levels on T-cells, B-cells, natural killer (NK) cells, and NKT cells from 26 RRMS patients under up to 72 weeks of therapy. Four-weekly infusions of natalizumab resulted in a significant and sustained increase of lymphocyte-bound natalizumab (p<0.001) which was paralleled by a significant decrease in detectability of the α(4) integrin subunit on all lymphocyte subsets (p<0.001). We observed pronounced natalizumab accumulations on T and B cells at single measurements in all patients who reported clinical disease activity (n=4). The natalizumab binding capacity of in vitro saturated lymphocytes collected during therapy was strongly diminished compared to treatment-naive cells indicating a therapy-induced reduction of α(4). Summing up, this pilot study shows that flow cytometry is a useful method to monitor natalizumab binding to lymphocytes from RRMS patients under therapy. Investigating natalizumab binding provides an opportunity to evaluate the molecular level of effectiveness of natalizumab therapy in individual patients. In combination with natalizumab saturation experiments, it possibly even provides a means of studying the feasability of patient-tailored infusion intervals. A routine-qualified biomarker on the basis of individual natalizumab saturation on lymphocyte subsets might be an effective tool to improve treatment safety.

Figure 5.2: The MS Ontology and its application: A) A basic formal ontology integration of MS Ontology and hierarchy of the concepts; B) Main classes of the MS Ontology and number of documents retrieved after mining PubMed via SCAIView; C) Tagged document after the integration of ontology in the search engine, SCAIView.

Figure 5.2 shows the hierarchy, main classes, and one application scenario of the ontology. The knowledge retrieved after integrating ontology into the semantic search engine SCAIView was used to model MS interaction map. The integration facilitated to find contextual knowledge of molecules and their role in disease mechanisms.

The enrichment of the ontology was done by using KNIME workflow querying UMLS database. In the following section the process has been discussed in details.

5.1.2 Automated data retrieval from UMLS:

The goal of this task is to enrich MSO with metadata retrieved from UMLS and to have an automated approach to retrieve the synonyms, definitions, and CUIs of the concepts. The retrieved data could then be embedded easily into the ontology by using OntoFast [208], which enriches the ontology in an automated manner. The proposed solution saves time by allowing users to retrieve data of many concepts automatically instead of doing a manual search. The starting point to enrich the ontology is to compile a list of concepts. Concepts from ontology were used to work further, there were more than one thousand concepts (1,170), and retrieving their attributes from UMLS database one at a time (or together) in one SQL query would have taken too long, as there could be hundreds of synonyms for some concepts. For example, the concept “4 Aminopyridine” has 306 synonyms with the language selected as “English” (Table 5. 1).

CUI	String	Relation	Synonym	Language	Definition
C0000477	4 Aminopyridine	SY	4 Aminopyridine	ENG	Potassium channel blocker,..
C0000477	4 Aminopyridine	SY	4-Aminopyridine	ENG	Potassium channel blocker,..
C0000477	4 Aminopyridine	SY	Pymadine	ENG	Potassium channel blocker,..
C0000477	4 Aminopyridine	SY	4-Pyridinamine	ENG	Potassium channel blocker,..
C0000477	4 Aminopyridine	SY	Aminopyridine 04	ENG	Potassium channel blocker,..
C0000477	4 Aminopyridine	SY	Fampridine	ENG	Potassium channel blocker,..
C0000477	4 Aminopyridine	SY	4-AP	ENG	Potassium channel blocker,..

Table 5.1: Some of the synonyms associated with the concept “4 Aminopyridine” retrieved from UMLS with the language selected English. Concept Unique Identifier (CUI), relation, and definition of the concept are also shown.

For the automated task KNIME was chosen. KNIME provides an easy way to automate these kinds of repetitive tasks, and there are ready-to-use nodes present to retrieve data from different databases. To avoid both performance issues with a large SQL query and repetitive manual work, a KNIME workflow has been built. The workflow reads each cell in a Microsoft Excel spreadsheet, queries the database, stores the retrieved results, and then reads the next cell of the Microsoft Excel file. The SQL query of “Database Reader” was modified to retrieve results based on a variable of the loop. The following terms were queried: CUI, definition of the concept, relation type (in our case Synonym), and language of the concept. The Synonym column brought all the variants of the original term, indicated by the abbreviation SY. The output of the workflow was a CSV file with different columns containing the queried terms (String), CUIs, definitions, relation types and synonyms. As depicted in Table 5.1 for the query “4 Aminopyridine”, only the values in the Synonym column of the output table changed. Since the relation queried was “SY” (Synonym in UMLS syntax), all the synonyms which had “4 Aminopyridine” as a heading term and had a relation type SY were retrieved. Repetitive values in columns Term, CUI, Definition, Relation, and Language indicated the attributes of the respective value in the SY field.

5.1.2.1 Evaluation of the system output

Some concepts were too vague to obtain an accurate retrieval; for example, the retrieved results of DC could be either Dendritic cell or Washington D.C., indicating that the queried terms should be clearer. In addition, we also noticed that querying with the biological entities i.e. genes or protein names generated two different results. One was associated with the gene whereas the other was associated with the protein name. Due to this reason, the output needed to be filtered depending upon the application scenario of the ontology. Besides that, phonotypical terms such as action potential, neurodegeneration, magnetic resonance imaging or drug names e.g. 4-aminopyridine were non-redundant. Although not all the ontology concepts were present in UMLS, most of the retrieved results for the concepts present in UMLS were correct according to the manual evaluation.

5.1.3 Automated Ontology Translation from UMLS:

By using a similar approach as described above, Spanish synonymous terms were also extracted and integrated in the ontology. The ontology was then able to extract knowledge from the Spanish text corpus. The semi-automated approach facilitated quicker performance of the task; in relatively short time we had an equally rich ontology as existed in the native language. It was observed that some molecules' names did not change, as the molecule names were based on different scientific findings. In addition, it was also found that multiple words terms (such as experimental autoimmune encephalomyelitis) were difficult to deal with, and that is why it is very important to assign CUIs manually before running the translation query. The Spanish terms enriched ontology has been used extensively to retrieve data from the Spanish EHR corpus (Table 5.2). A similar task can be performed with MeSH terms, and any concept can be translated into one or all of the 16 available languages of MeSH terms. The system not only works for UMLS, but it is also possible to retrieve data from any other database as the correct mapping of identifiers is important.

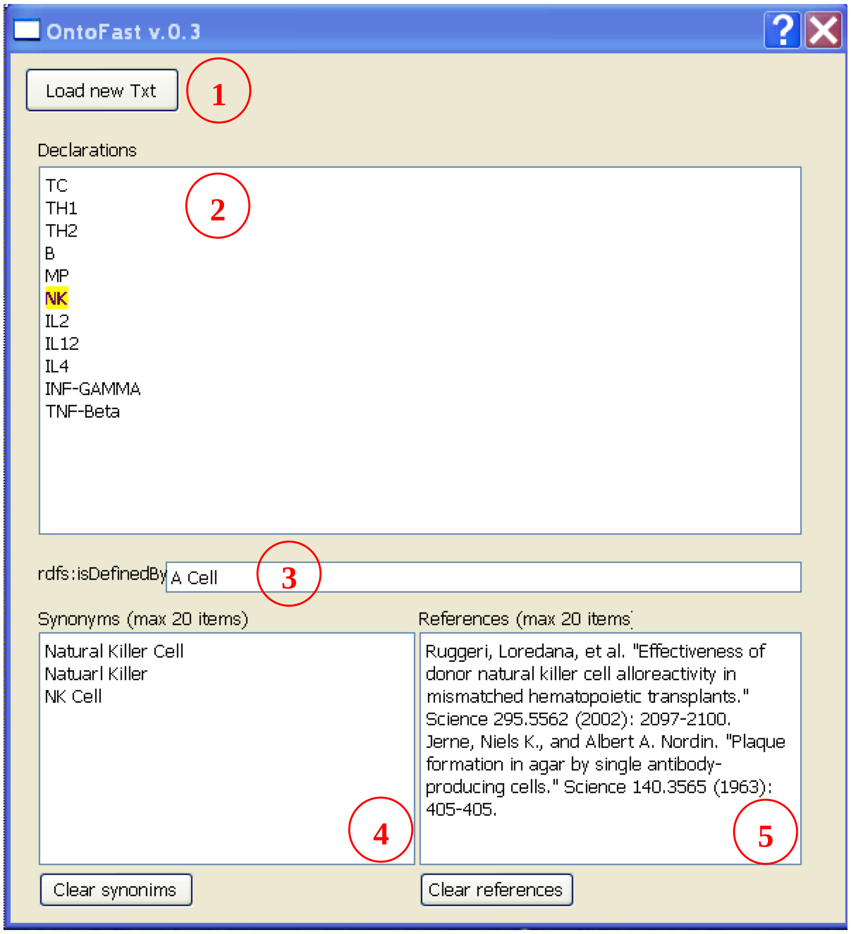
English Terms	CUI	Spanish Terms	Language
Acetylcholine	C0001041	Acetilcolina	SPA
Esophageal Reflux	C0017168	Acid reflux	SPA
Homocysteine	C0019878	Acido 2-Amino-4-mercaptobutirico	SPA
4-Aminobutyric Acid	C0016904	Acido 4-Aminobutirico	SPA
Folic Acid	C0016410	Acido Folico	SPA
Acid Synthase, Fatty	C0015683	Acido Graso Sintasa	SPA
5-HIAA	C0020361	Acido Hidroxiindolacetico	SPA
Actin	C0001271	Actina-alfa	SPA
Activation, Lymphocyte	C0024262	Activacion de Linfocitos	SPA
Behavior, Sex	C0036864	Actividad Sexual	SPA
Hypoesthesia	C0020580	Adormecimiento	SPA
Adrenaline	C0014563	Adrenalina	SPA
Aphasia	C0003537	Afasia	SPA
Agents, Anticholinergic	C0242896	Agentes Anticolinergicos	SPA
Agents, Antidepressive	C0003289	Agentes Antidepresivos	SPA

Table 5.2: The resulting table after the translation query execution. English term, identifiers, Spanish Terms and Language columns are shown. In UMLS syntax, SPA represents Spanish. The important aspect of the retrieval was to provide the correct identifier and language of output.

A recently developed tool (OntoFast, discussed below) [208] facilitates ontology enrichment and it required minimum efforts to add all the Spanish concepts into a pre-existing ontology. It provides an easy-to-use interface and can take a list of hundreds of concepts in one go. Relations addition is not included in OntoFast, as the aim of this program development is to have an enriched ontology with synonyms, references, and definition. It was not meant to provide relations of the concepts as mapping relations is a trivial scenario for any ontology editing program e.g. Protégé.

5.1.4 OntoFast:

In this section we discuss about an application, OntoFast [208], which allows to speed up the standard procedure of ontology development and metadata integration. Usually these processes take anywhere between many months to a couple of years and involves many people. For example, Protein-Ligand Interaction Ontology (PLIO) [264] was developed in 18 months, Multiple Sclerosis Ontology (MSO) [28] was developed in one year and Gene Ontology (GO) took many years and is still being updated. One of the main hurdles while developing MSO was the difficulty of introducing new concepts into the Protégé user interface. This task proved to be time consuming and labor intensive. Since the ontology engineers are specialists in their domain and they may develop ontologies only for their specific needs, they are usually not experts in ontology development work. This lack of practice often slows down the progression of the work and forces them to do repetitive tasks which can be automated easily. OntoFast solves this problem by providing an easy-to-use and convenient interface which can prevent domain experts wasting the precious time. More than one synonym and reference can be given in different lines by copy paste, making it more convenient for information retrieval systems to broaden the coverage of the ontology. Different options allow users to embed definitions, synonyms and references of the ontology via an easy-to-use graphic user interface. Since ontologies can be designed with different hierarchies and different application scenarios which vary from domain to domain and from task to task, hierarchical feature were not added into it. The output of the program can be easily opened with any standard ontology editor like Protégé. Then the hierarchy can be customized by simple drag and drop, according to the user’s specific needs. Figure 5.3 shows the interface of the application with different options, a button to load a Text file of concepts, a field to add definition, and text boxes for synonyms and references.



Declarations

TC
TH1
TH2
B
MP
NK
IL2
IL12
IL4
INF-GAMMA
TNF-Beta

rdfs:isDefinedBy: A Cell

Synonyms (max 20 items)

Natural Killer Cell
Natuarl Killer
NK Cell

References (max 20 items)

Ruggeri, Loredana, et al. "Effectiveness of donor natural killer cell alloreactivity in mismatched hematopoietic transplants." Science 295.5562 (2002): 2097-2100.
Jerne, Niels K., and Albert A. Nordin. "Plaque formation in agar by single antibody-producing cells." Science 140.3565 (1963): 405-405.

Figure 5.3: OntoFast interface and description of various options 1) Load button for loading a list of concepts from a txt file. 2) Concepts list which shows concepts loaded from a txt file. A concept can be selected by clicking on it. The selected concept will be highlighted. 3), 4) and 5) fields are for defining basic properties of the selected concept. Synonyms and References fields can take more than one value.

In earlier section of the current chapter, it has been shown that metadata associated with biomedical concepts can be retrieved automatically from UMLS by using KNIME workflow [29]. In the following section, the approach to build an ontology quickly using OntoFast, from the metadata retrieved, will be discussed.

Importing concepts in OntoFast is very easy, since the list of concepts can be imported by clicking on the “Load new Txt” button. All of the concepts of a prospective ontology can be given in the form of a list in a text file (.txt). Fields in the text file should be separated by carriage return commands. The application reads each new line as a new concept and generates the list of concepts that is visualized in the “Declarations” text box. The associated metadata can be then added by selecting a concept in the list. Selected concepts will be highlighted. Just after importing the list of declarations, the application asks to choose the output .xml file, which can be used in Protégé or in any other ontology editing application. From this moment, the user will not need to take care of manually saving the output .xml file, since the application will execute automatic saving every time a different concept is selected, as well as on exit.

Metadata can be easily associated with the imported concepts by selecting a concept and providing the associated details in relative fields. The main attributes required for the ontology were definitions, synonyms and references; thus there are different text boxes given to incorporate the same attributes. As the goal was to speed up the initial step in the development of new ontologies, each of the boxes can accommodate copy/paste to quickly populate the ontology. In addition, more than one synonym and reference can be given in different lines. Finally, the hierarchy of the ontology can be arranged later on by the user in Protégé, since such an operation can be carried out very quickly within it. OntoFast can be downloaded from http://www.francescopappalardo.net/ontofast.zip.

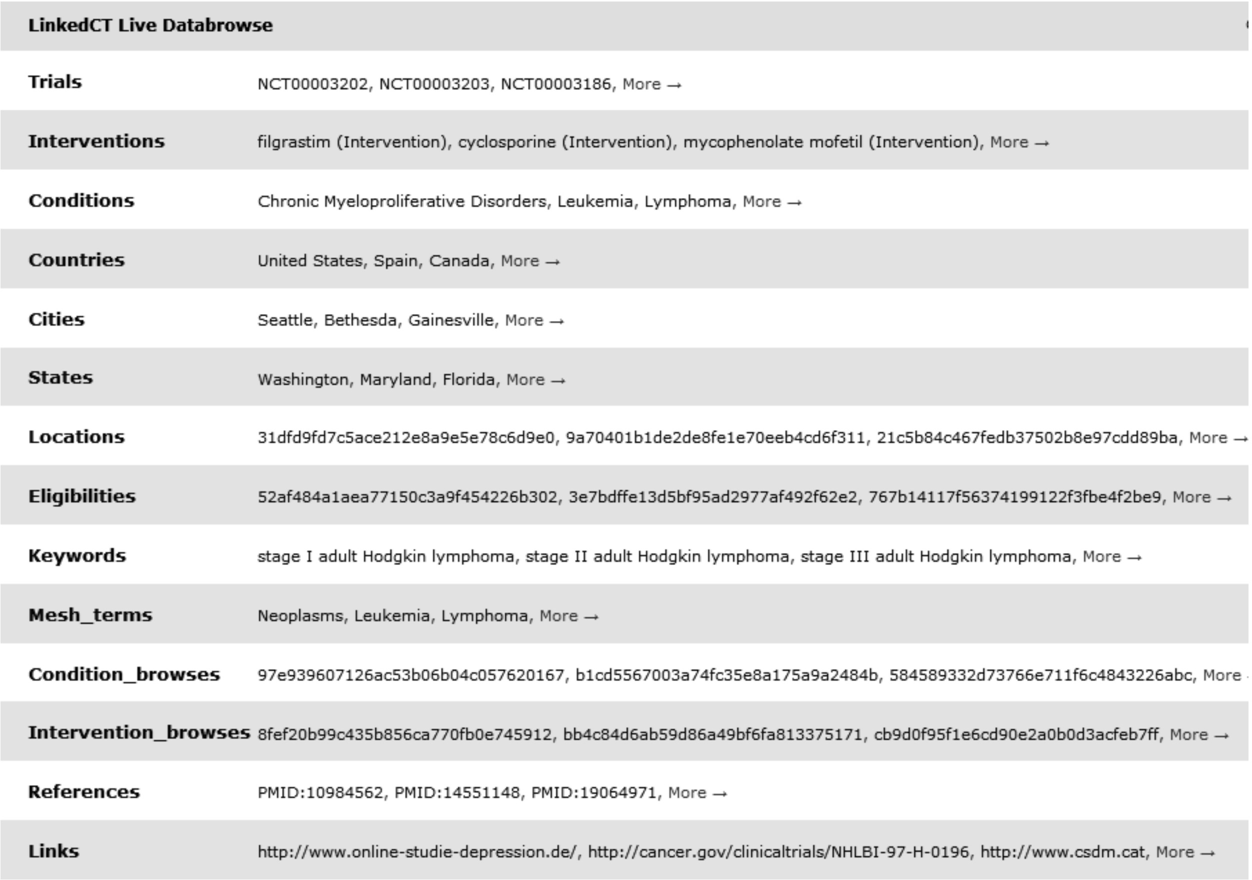
After retrieving the data from a biomedical database and developing the disease ontology, the next step was to gather publically available data e.g. Clinical trials etc. An innovative approach has been used to download a clean dataset from public source.

5.1.5 Corpus generation from web to analysis the content and applying NLP tools:

The approach described here allows easy and efficient storing of web pages, as well as generates a clean corpus of relevant data. The requirement of this approach is as follows.

The corpus generation methodology was developed to facilitate the web mining and make it easier to get a clean corpus in no time. There are many web mining tools dedicated to perform this job with different levels of complexity, including depth of the weblinks as well as parsing in real time while crawling the web. These tools, however, do not have a straightforward and simple GUI to perform the simple task of gathering a clean dataset. Nutch [265], Websphinx [266], and openwebspider [267] are some of the popular tools available to do this job. That said, extracting and getting corpus straight from web-links is such a basic task which requires neither such advanced tools nor the skills to operate them. Besides, the output of these tools may contain irrelevant files & folders and a not-so-clean dataset. In addition, the requirement to have a clean corpus pushed us to have a methodology which could be used on any site with minimal modification and contained the highest percentage of content purity. The other tools mentioned above crawl the web, thus users have limited control on the content which is downloaded, gathered, extracted, or retrieved. They do not offer the best option to have a clean dataset. Our methodology retrieves a clean dataset from any given website (with a few limitations) and the resulting dataset would only contain the data required by the user. The dataset then could be further used to perform different linguistic tasks to get the best results.

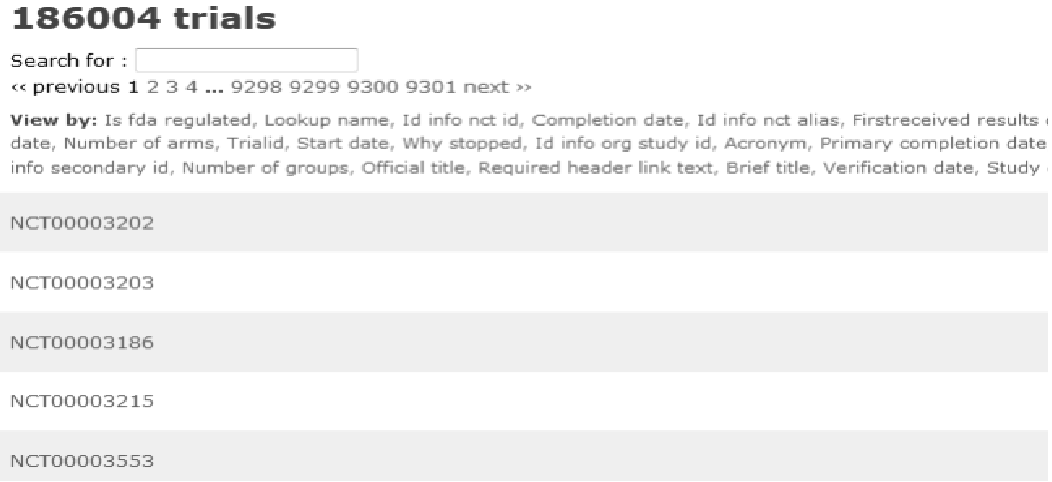
Use case: Here we present our use case with the website www.LinkedCT.org and how we processed the site with our methodology, resulting in a clean dataset of clinical trials specific for RRMS. LinkedCT is a semantically processed site of www.clinicaltrials.gov (Figure 5.4).



LinkedCT Live Databrowse	
Trials	NCT00003202, NCT00003203, NCT00003186, More →
Interventions	filgrastim (Intervention), cyclosporine (Intervention), mycophenolate mofetil (Intervention), More →
Conditions	Chronic Myeloproliferative Disorders, Leukemia, Lymphoma, More →
Countries	United States, Spain, Canada, More →
Cities	Seattle, Bethesda, Gainesville, More →
States	Washington, Maryland, Florida, More →
Locations	31dfd9fd7c5ace212e8a9e5e78c6d9e0, 9a70401b1de2de8fe1e70eeb4cd6f311, 21c5b84c467fedb37502b8e97cdd89ba, More →
Eligibilities	52af484a1aea77150c3a9f454226b302, 3e7bdffe13d5bf95ad2977af492f62e2, 767b14117f56374199122f3fbe4f2be9, More →
Keywords	stage I adult Hodgkin lymphoma, stage II adult Hodgkin lymphoma, stage III adult Hodgkin lymphoma, More →
Mesh_terms	Neoplasms, Leukemia, Lymphoma, More →
Condition_browses	97e939607126ac53b06b04c057620167, b1cd5567003a74fc35e8a175a9a2484b, 584589332d73766e711f6c4843226abc, More
Intervention_browses	8fef20b99c435b856ca770fb0e745912, bb4c84d6ab59d86a49bf6fa813375171, cb9d0f95f1e6cd90e2a0b0d3acfeb7ff, More →
References	PMID:10984562, PMID:14551148, PMID:19064971, More →
Links	http://www.online-studie-depression.de/, http://cancer.gov/clinicaltrials/NHLBI-97-H-0196, http://www.csdm.cat, More →

Figure 5.4: Home page of LinkedCT [268], clinical trials are categorized by different headers.

LinkedCT provides categorization of clinical trials based on clinical trials ID, Interventions, conditions, locations, MeSH terms, and reference of the publications. This classification makes it easier to find any clinical trial from its metadata. There were approximately 1,86,004 processed clinical trials at the time of this writing (Figure 5.5).



186004 trials

Search for :

<< previous 1 2 3 4 ... 9298 9299 9300 9301 next >>

View by: Is fda regulated, Lookup name, Id info nct id, Completion date, Id info nct alias, Firstreceived results date, Number of arms, Trialid, Start date, Why stopped, Id info org study id, Acronym, Primary completion date info secondary id, Number of groups, Official title, Required header link text, Brief title, Verification date, Study

NCT00003202

NCT00003203

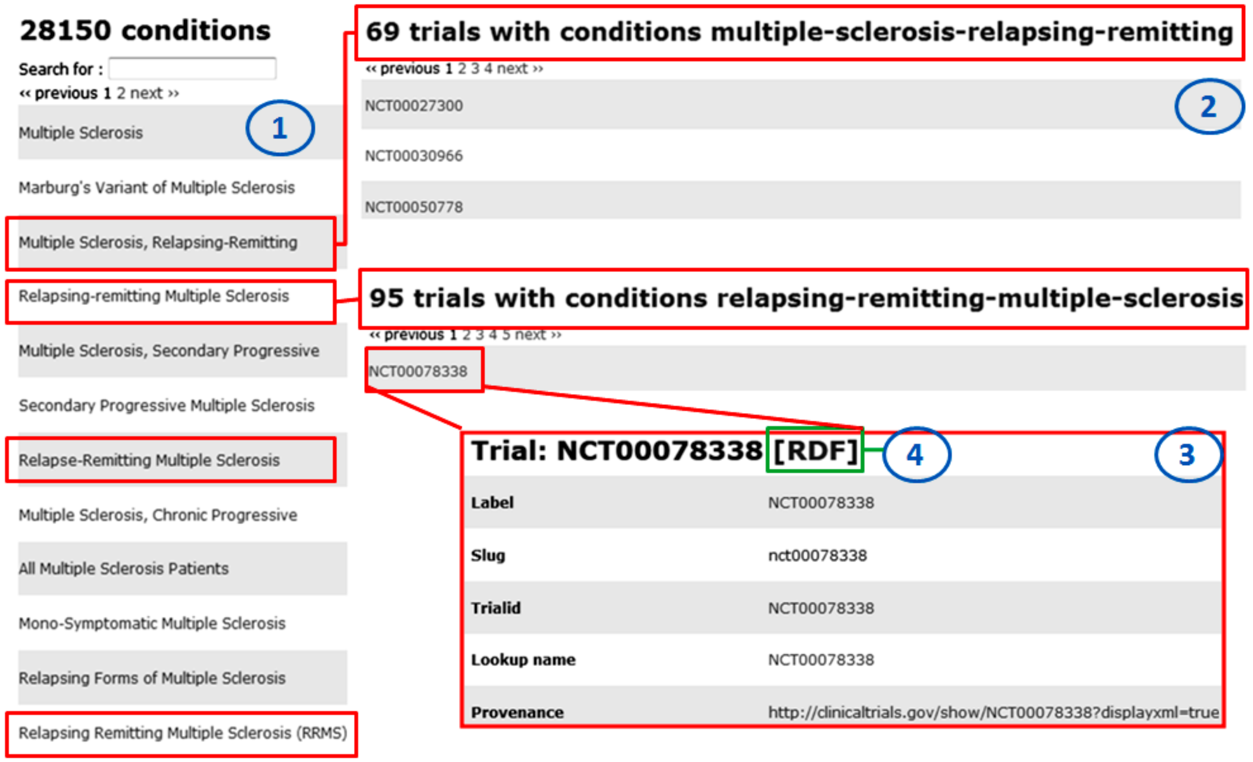
NCT00003186

NCT00003215

NCT00003553

Figure 5.5: Total number of clinical trials in Linked CT.org.

The requirement of the work is to download all the clinical trials associated with a particular disease subtype (RRMS) and those clinical trials were stored beneath four variations of disease name (Multiple sclerosis Relapsing-Remitting, Relapsing-remitting Multiple Sclerosis, Relapse-Remitting Multiple Sclerosis, and Relapsing Remitting Multiple Sclerosis). The actual data files (RDFs) were stored beneath three html pages of each of the disease name label (Figure 5.6(1)). We stored each of the page under disease name as html file and then merge them together to have all the NCT trial numbers on one html page (though they were stored with the variations of disease name on the website) (Figure 5.6(2)). It was found that the pattern of RDF data file URL and the page where it contains the link of the RDF data file (Figure 5.6(3)) doesn't differ much and there is a similar pattern for each RDF data file URL (Figure 5.6(4)) associated with the webpage link. Further, we extracted all the links by using LinkGopher from the merged NCT trial numbers page and then observed at the patterns of RDF data file URLs and html pages URLs. After finding out the pattern, keywords have been replaced with the one which was associated with RDF data files and then all the RDF files were downloaded by using DownThemALL.



28150 conditions

Search for :

« previous 1 2 next »

Multiple Sclerosis

Marburg's Variant of Multiple Sclerosis

Multiple Sclerosis, Relapsing-Remitting

Relapsing-remitting Multiple Sclerosis

Multiple Sclerosis, Secondary Progressive

Secondary Progressive Multiple Sclerosis

Relapse-Remitting Multiple Sclerosis

Multiple Sclerosis, Chronic Progressive

All Multiple Sclerosis Patients

Mono-Symptomatic Multiple Sclerosis

Relapsing Forms of Multiple Sclerosis

Relapsing Remitting Multiple Sclerosis (RRMS)

69 trials with conditions multiple-sclerosis-relapsing-remitting

« previous 1 2 3 4 next »

NCT00027300

NCT00030966

NCT00050778

95 trials with conditions relapsing-remitting-multiple-sclerosis

« previous 1 2 3 4 5 next »

NCT00078338

Trial: NCT00078338 [RDF]

Label	NCT00078338
Slug	nct00078338
Trialid	NCT00078338
Lookup name	NCT00078338
Provenance	http://clinicaltrials.gov/show/NCT00078338?displayxml=true

Figure 5.6: The layers of web pages and actual dataset require for the work (marked with number 4 and labeled with RDF).

5.1.6 SBML2SMW, Transforming Systems Biology knowledge into MediaWiki Pages:

One of the important aspects of our work was to preserve and reuse the knowledge extracted from different sources so it can be used for Systems biology modeling work. For this, MediaWiki software was chosen as the central component of the system due to its interoperability and reliability. In addition, there are many off-the-shelf plug-ins available which extend the functionality of MediaWiki. However, there was no Systems biology connection to Semantic MediaWiki and to fill this gap; we aimed to have a framework which would have different capabilities (discussed below in details).

5.1.6.1 User Requirements: Interface and infrastructure:

An easy-to-use graphical user interface was a prerequisite, since the primary users of the system would be biologists. Due to the collaborative work and enriching the knowledge base, it was important that each user's activity could be appropriately monitored and an administrator must be able to trace and if needed, revert individual changes. Data retrieval has to be easy and all the associated knowledge with any particular entity should be available with a few clicks. Interoperability with other systems via open APIs is also required. Portability of the back-end must not be an issue and it should work on any standard Windows/Linux servers and on any portable computing device (from PC to Mac and Tablets).

5.1.6.2 Modeling software as front end:

The modeling software must support SBML and the graphical user interface should be rich enough to represent biochemical networks and health related issues (e.g. adverse drug reactions and drug-drug interactions). It should be able to connect with external databases and have robust capabilities of network analysis and quantitative modeling.

5.1.6.3 Knowledge base as back end:

The semantics of the knowledge base should be explicit, completely ontologically annotated, and interoperable with different knowledge sources. The knowledge base must be stable and have a proven record of flexibility and scalability. A state-of-the art and open source solution is needed.

After thorough research and market analysis, it has been concluded that all requirements described above could be addressed by combination of two publically available tools, CellDesigner and Semantic MediaWiki (SMW). The development of a "semantic glue" (to bridge these two technologies) is the best option available to solve the issue. Hereafter the reasons for the choice in the context of the requirements have been described.

5.1.6.4 Evaluation of off-the-shelf technologies:

Several implementations of wiki exist in biology (e.g. WikiGenes [269], WikiProteins [270], and WikiPathways [271], unfortunately none of them matched the requirements. For example, WikiPathways enables community curation; however, it does not enable dynamically importing connections as found in other pathways stored in WikiPathways, since the pathways are stored as "In-silos" (each separates from the others). Another example is the Payao system [239], which enables a more systematic, community-based annotation and curation with SBML and SBGN compliance; but they are "network centric" and thus do not share the relations between pathways. It has also been shown [272] that Semantic MediaWiki has the capabilities to write labeled links to create RDF triples. It is very simple to use and can serve as a useful tool for collaborative editing according to simple RDF statements, which is also appropriate for biology. More recently, the project HALO has shown [273] that low-cost highly-scalable modeling of basic scientific knowledge in health science could be appropriately handled with Semantic MediaWiki and a further added biology-friendly extension (SMW+). Therefore, MediaWiki and the SMW extension have been chosen as the core technology for the Systems biology knowledgebase. In order to enable CellDesigner to access (in read/write mode) the SMW, the SBML2SMW plug-in [274] has been used. That plug-in has been developed independently and at the same time as the PathwayAccess [275]. It does not exploit their API, but in order to bridge the internal data representation of CellDesigner and the SMW, it exploits a minimal ontology and an ontology mapping service. This makes it on-the-fly compatible with any RDF environment e.g. IWB [276]. The current infrastructure enables a variety of scenarios, including: collaborative knowledge acquisition and hypothesis generation, automatic knowledge update via text mining, knowledge condensation (e.g. using CellDesigner and PathwayAccess plug-ins), and large scale knowledge reasoning (e.g. exposing the content with a SPARQL entry point).

5.1.6.5 The Integrated framework and Solution: CellDesigner: Advanced front-end for systems biology domain experts:

The features of CellDesigner and its various functionalities have been discussed in chapter 4.

5.1.6.6 SMW: Front-end for occasional users and powerful knowledge storage:

SMW is a free extension of MediaWiki that adds semantic annotations, therefore allowing the wiki to function as a collaborative database with semantically tagged content. MediaWiki has a large variety of extensions which make it directly pluggable into the semantic web as a SPARQL entry point and as a linked data server. In view of the very solid semantic foundation of SMW, OWL DL (Web Ontology Language, Description logic), and its proven robust wiki structure relying on MediaWiki (the software behind Wikipedia.org), it has been decided to use it as the core technology. The RDF export capabilities of SMW allow the seamless transfer of its content to other powerful semantic stores. To use SMW as core technology, the only problem to be solved was to create a bidirectional bridge between SMW and CellDesigner.

5.1.6.7 CellDesigner - SMW integration:

To integrate these two different platforms; a plug-in of CellDesigner SBML2SMW has been developed, which transforms all the knowledge curated in CellDesigner model into Mediawiki knowledge base. This facilitates:

- Acquisition of semantically enriched and scattered biomedical knowledge from different sources (i.e. DBGET, SGD, iHOP, Genome Network Platform, PubMed, Entrez Gene, SABIO-RK) as shown in figure 5.7.
- Sharing and reusing knowledge networks in the context of biomedical hypotheses generation

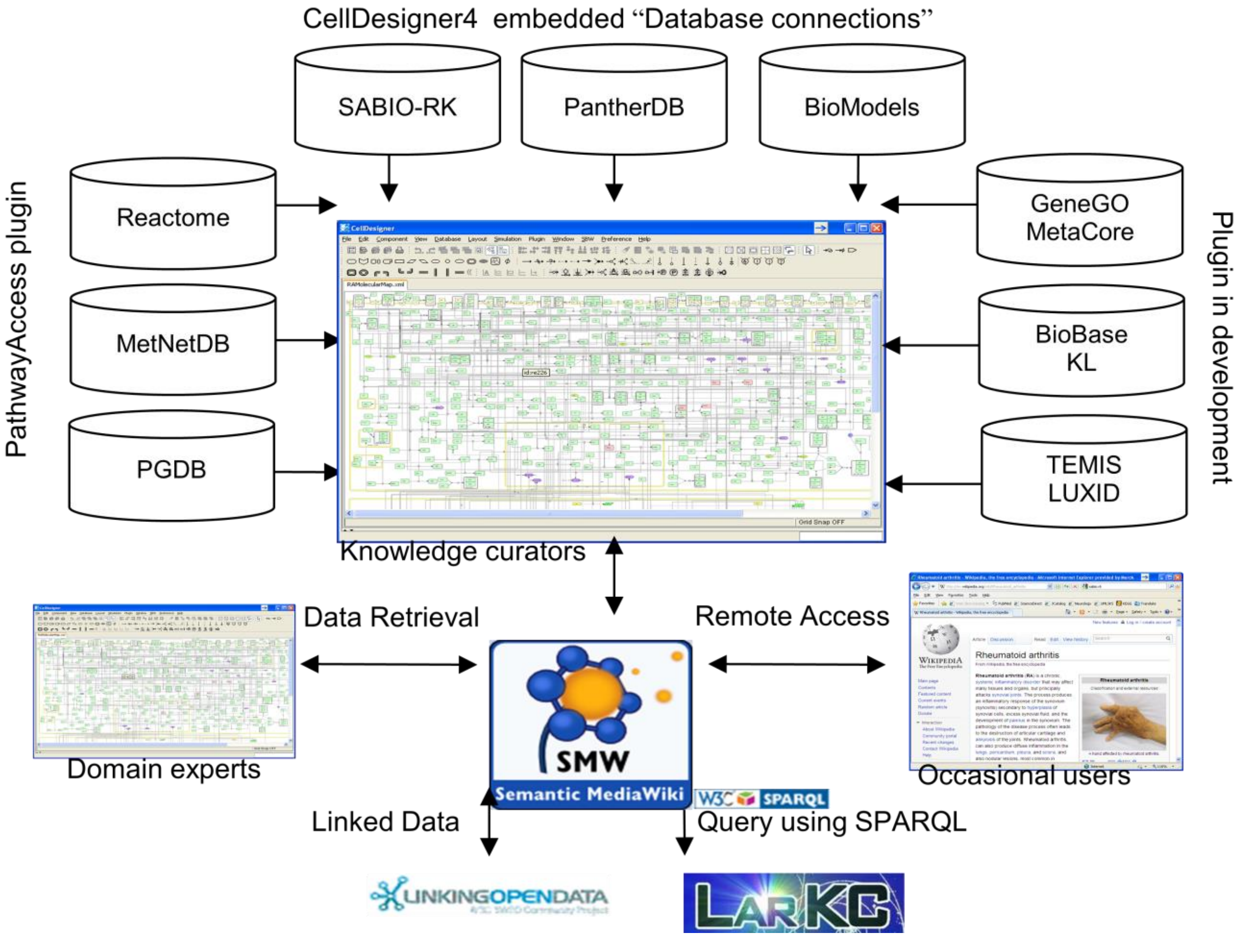


Figure 5.7: The overall workflow of the system. CellDesigner retrieves data from different databases and with the tool discussed (SBML2SMW), it can convert the data into web pages, where a user can read it through web browser or an expert can get in-depth knowledge from the CellDesigner application. External queries can also be executed by using different applications like LarkC etc.

CellDesigner imports data from different knowledge sources and with the SBML2SMW plug-in, it can convert the data into semantic wiki pages. At this point, an occasional user can view it, comment on it, and edit it through a web browser; an expert can perform quantitative modeling using SBW or other tools e.g. Copasi [277] etc. External queries can be executed by using different applications like LarkC [278]. The semantic knowledge base can be automatically enriched by different text mining tools (e.g. ProMiner, Luxid, GATE, BioNotate, I2E, NCBOAnnotator, etc.), given that these tools are able to generate RDF using a common set of ontologies. In addition to the SBML2SMW plug-in, a "translation server" has to be deployed on the client or on the server, which automatically maps the curated models between the core ontologies used in CellDesigner and the SMW model. This ontology mapping has been modeled in OWL DL for coherence with the SMW model. It was intentionally kept small and concise, covering only the required aspects of SBML models, and can be easily extended. A graphical representation of the resulting ontology is shown in Figure 5.8.

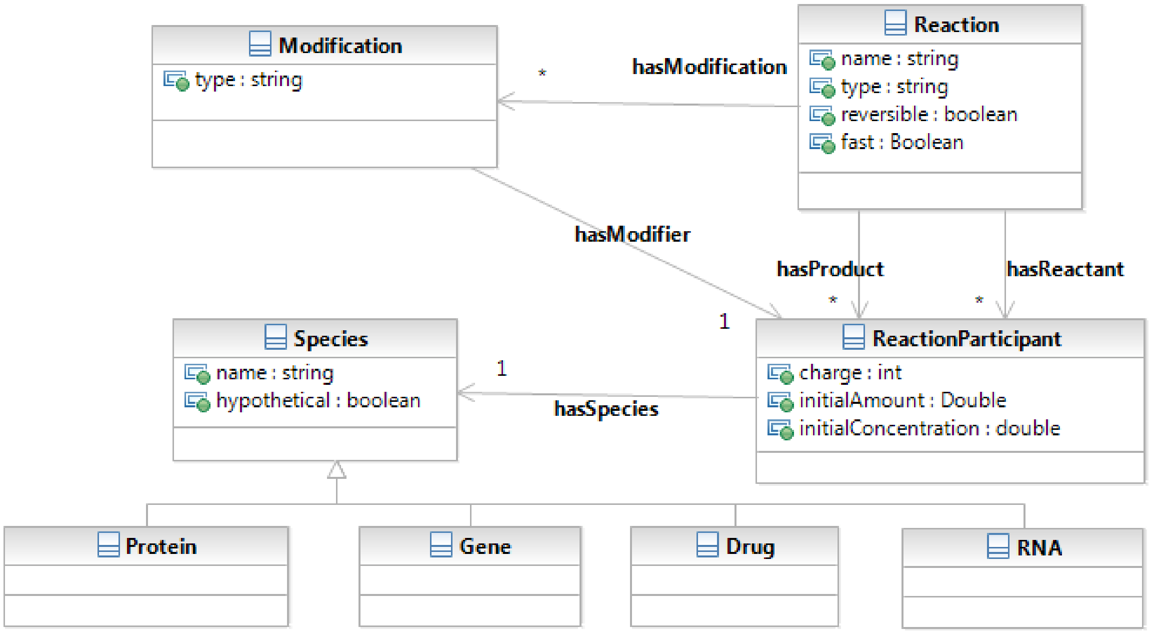


Figure 5.8: A mapping ontology between CellDesigner models and Semantic MediaWiki.

After all the foundation work of information representation, in the next section, we will discuss about the approaches which have been used to model MS.

5.2 Modeling of MS disease:

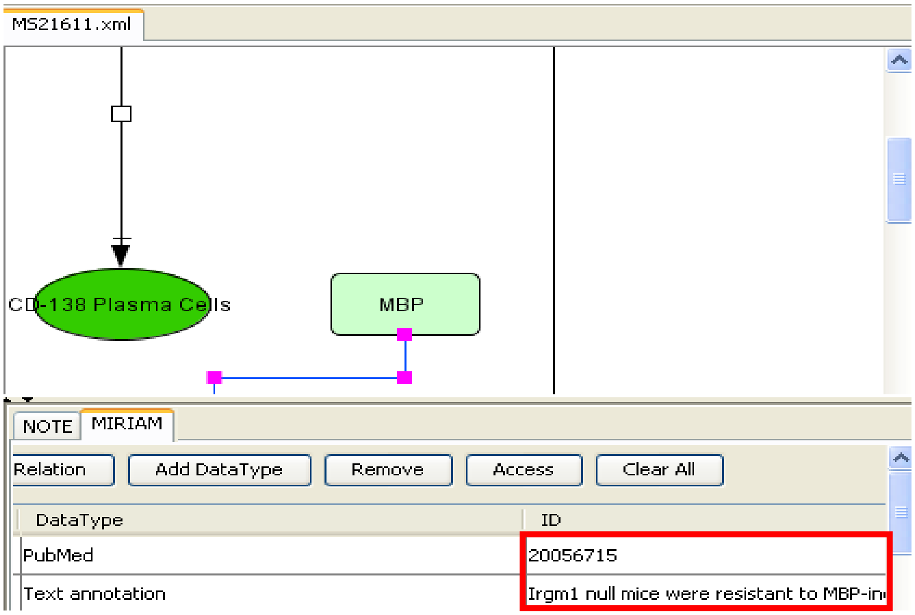
5.2.1 Molecular interaction map of MS:

The molecular interaction map of MS was developed to reveal the underlying mechanisms of the disease by laying interacting molecules and pathways associated with them. A molecular interaction map reveals the mechanisms of different pathways; and by looking at it, one can instantly become familiar with the mechanism of different drugs and their interactions with other molecules. The clinical trials overlay on molecular interaction map also provides a means to see the clinical phenotype caused by molecular interactions. DrugBank data provides a detailed description about drug behavior. The MS molecular interaction map was developed after the evaluation of different commercial suites i.e. Genego, IPA, and Biobase. None of those software suites provided the function of sentence support underneath the edge. The Miriam section of CellDesigner was extended to accommodate and store the textual support of the relation.

5.2.1.1 Disease model development and validation:

The model was split into 3 parts due to its large size, as it contains ~650 nodes and ~900 connections. Most of the nodes are proteins or bio-molecules and then phenotypes associated with them. There are also genes which show significant association with the disease. The map also contains drugs acting on different molecules, as clinical trials data was overlaid after the literature sources.

There were five interaction types, represented as lines with arrowheads and other shapes. Different shapes of links represent different interaction types e.g. inhibition, activation, increment, decrement and modulation. The legend set of edges was adopted due to the limited expressiveness of CellDesigner edges. For example, the default legend set in CellDesigner can assign an increment or decrement in a process but a protein expression cannot be changed (increment or decrement) by the actions of other biological molecules. However, in human body it does happen. To keep the model simple, a new legend set has been introduced which represents the actions of biological molecules on other set of molecules either increment or decrement. Under each edge, the supporting sentence was embedded with the PubMed ID of the scientific paper from which it was taken (Figure 5.9).



DataType	ID
PubMed	20056715
Text annotation	Irgm1 null mice were resistant to MBP-in

Figure 5.9: The supported sentence extracted from scientific literature, after manual reading, and PubMed ID stored under each edge in the molecular interaction map.

The following are the seven types of entities: genes, proteins, receptors, drugs, cells, phenotypes, and degradation. Aside from degradation and phenotype, all entities have their UMLS unique identifier stored within the interaction map. The names of the molecules were taken as mentioned in the literature, but later they were mapped as proper UMLS identifiers by searching each of them in UMLS database.

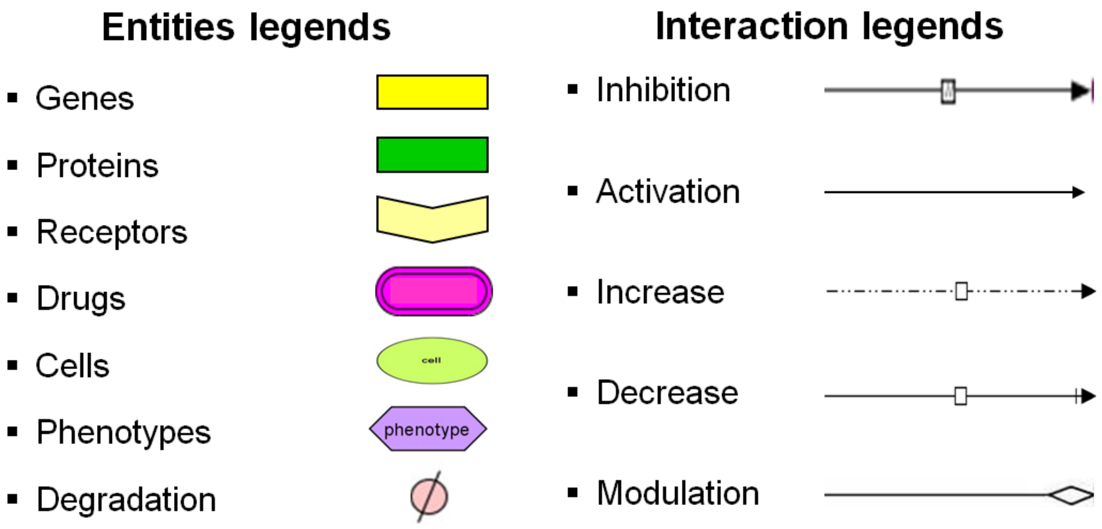


Figure 5.10: CellDesigner entities legends and interactions legends used to model MS map.

Following Knowledge set was embedded under each entity and Interaction:

1) UMLS CUI (Entities)
2) Sentences in Miriam (Interactions)
3) PubMed ID (Interactions)

The interaction map was one of the first of its kind, as it was developed by reading literature manually, and the support of each edge is available within the map. The map was also transformed into a complete wiki running internally with the help of the CellDesigner plug-in SBML2SMW [274]. In addition, the knowledge within the map was linked by LinkedOpenData consortium datasets by using an external application "Information workbench" [276]. This allowed us to dig deeper into each of the nodes and edges in real-time with transient retrieval of information by using federated queries to various datasets. Figure 5.11 gives a glimpse of the molecular interaction map of MS, since the map was large enough thus only a portion of map is shown here. The MS map has been published at Payao website [http://sblab.celldesigner.org/Payao10/bin/].

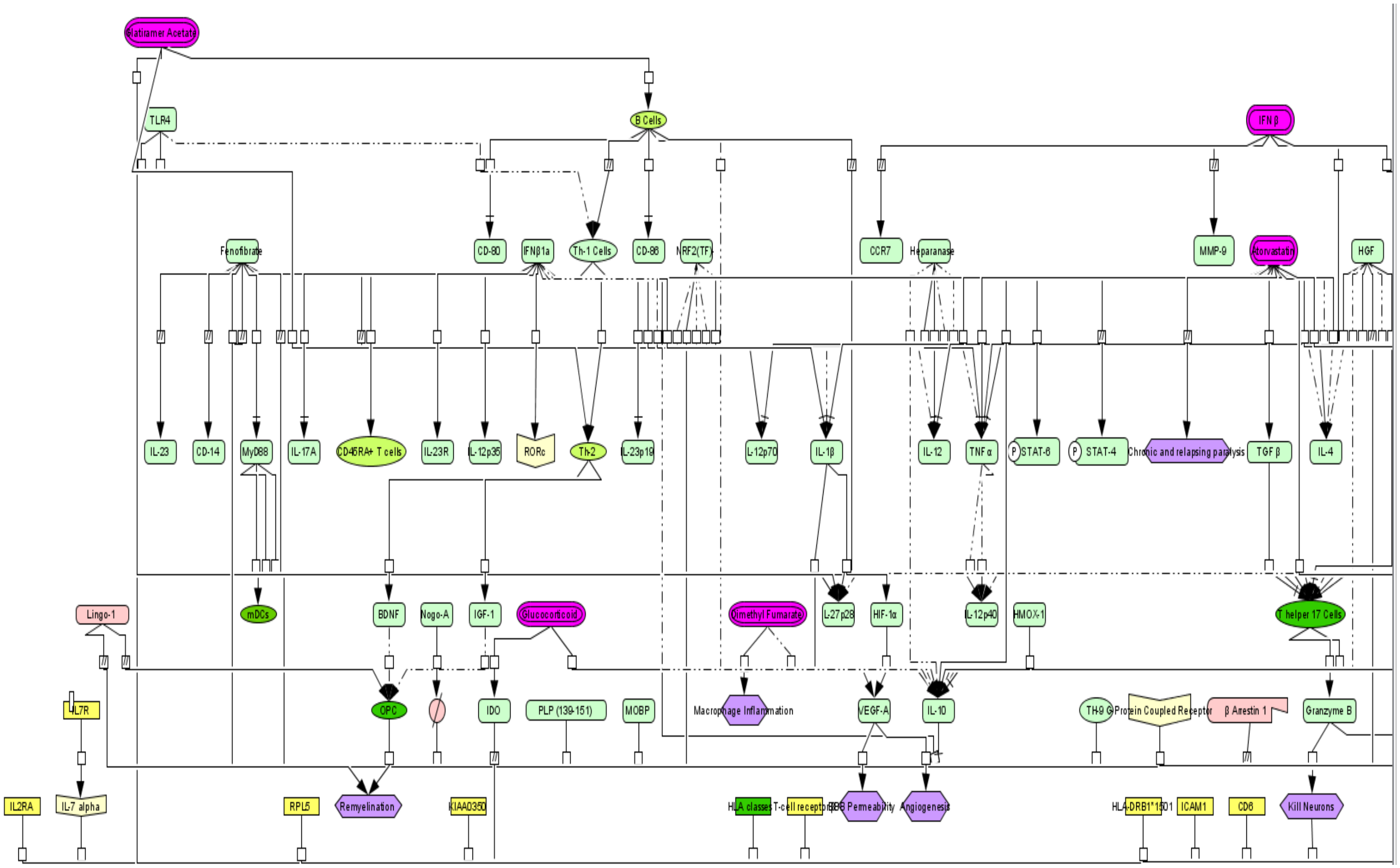
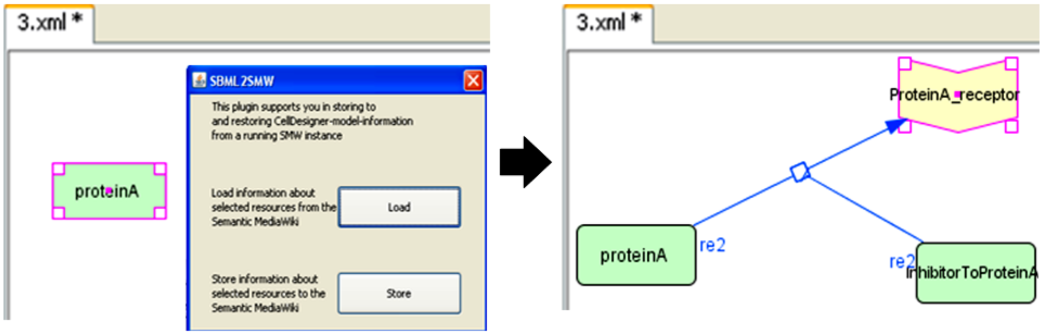


Figure 5.11 : One part of the MS Model. Model was split into three parts due to its large size.

5.2.1.2 Example usage of SBML2SMW with CellDesigner Model:

The architecture described in the previous sections opens up semantic web technologies to new user groups and applications. When a biologist models complex facts in CellDesigner, he or she can easily populate SMW with the newly discovered knowledge. Users can edit information directly in the wiki without the need of CellDesigner. Advanced users can also run complex SPARQL queries. For example, an expert user could easily find all reactions having a generic protein marked as hypothetical and catalyzed by catalyst X, by running one SPARQL query. The complexity of the query can be extended by combining this knowledge with other knowledge sources i.e. LinkedOpenData [279] , and using more powerful platforms e.g. LARKC [278].

Evaluation: The system has been used for the last 20 months, and it is smooth and stable. Initially, the system was capable of retrieving entities but not relations. A newer version of the plug-in (SBML2SMW) was developed with the feature of relations retrieval among the entities as well as all the data associated with them, such as supported sentences for the relations and PMID of the papers from where the supported sentence was taken. In addition, under each entity there are the UMLS concepts' identifiers which can also be retrieved from the SMW pages. The system is being used by our biologist colleagues at different geographical locations without any issue. Figure 5.12 shows the entity before the retrieval of the previously stored knowledge, and after retrieval, accomplished by selecting the entity (ProteinA) and pressing the "Load" button from the SBML2SMW popup box. All the associated knowledge with ProteinA will be retrieved from the backend SMW and displayed in CellDesigner.



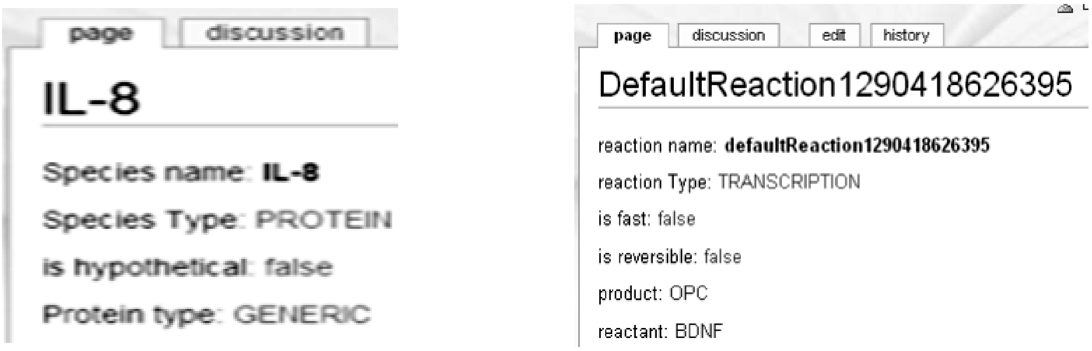
This plugin supports you in storing to and restoring CellDesigner-model-information from a running SMW instance

Load information about selected resources from the Semantic MediaWiki

Store information about selected resources to the Semantic MediaWiki

Figure 5.12: Retrieval of knowledge associated with an entity by using SBML2SMW plug-in. The newly added entities were stored previously by using store option of SBML2SMW plug-in into backend Semantic MediaWiki.

Before retrieval, a model has to be stored by using SBML2SMW plug-in store feature and SBML2SMW transformed this knowledge into backend SMW. After the transformation of the model, the knowledge curated within the model can be seen as wiki pages (Figure 5.13).



IL-8

Species name: **IL-8**

Species Type: PROTEIN

is hypothetical: false

Protein type: GENERIC

DefaultReaction1290418626395

reaction name: **defaultReaction1290418626395**

reaction Type: TRANSCRIPTION

is fast: false

is reversible: false

product: OPC

reactant: BDNF

Figure 5.13: MediaWiki pages of stored entity (left) and interaction (right) of MS map. On reaction page, type of reaction, name of product and reactant can be seen.

5.2.1.3 Overlaying Clinical Trials and DrugBank data:

Clinical trials and DrugBank data was overlaid on the interaction map to incorporate the phenotypical findings in the literature as well as in the clinical trials. DrugBank data was added to have drugs mode of action and molecules they target. The clincial dataset used was from NMSS, as described in the methodology section. The integration helped us to show a broader aspect of a disease by combining different streams of knowledge. In addition to clincial data, drugs' information was taken from the DrugBank RDF repository to run complex queries. The final integrated map contained molecular entities from literature, phenotypes from clinical trials and drug knowledge from DrugBank.To the author's knowledge, this is the first molecular interaction map to use these different streams of knowledge to represent a neurodegnerative disease.

In addition to overlaying the DrugBank data, we also looked at the modes of action and possible mechanisms of the drugs and created a chart based on the most frequently occuring disease mechanisms associated with MS disease in the clinical trial dataset. Figure 5.14 shows that most of the drugs being tested interact with a mechanism involving the immune system in order to treat the disease. This exercise has been done as some of the drugs for the treatment of MS have unknown functionalities, and eventhough a wide body of literature supports the notion that MS is primarily a disease of the immune system, there is still a disagreement within the scientific community over this theory.

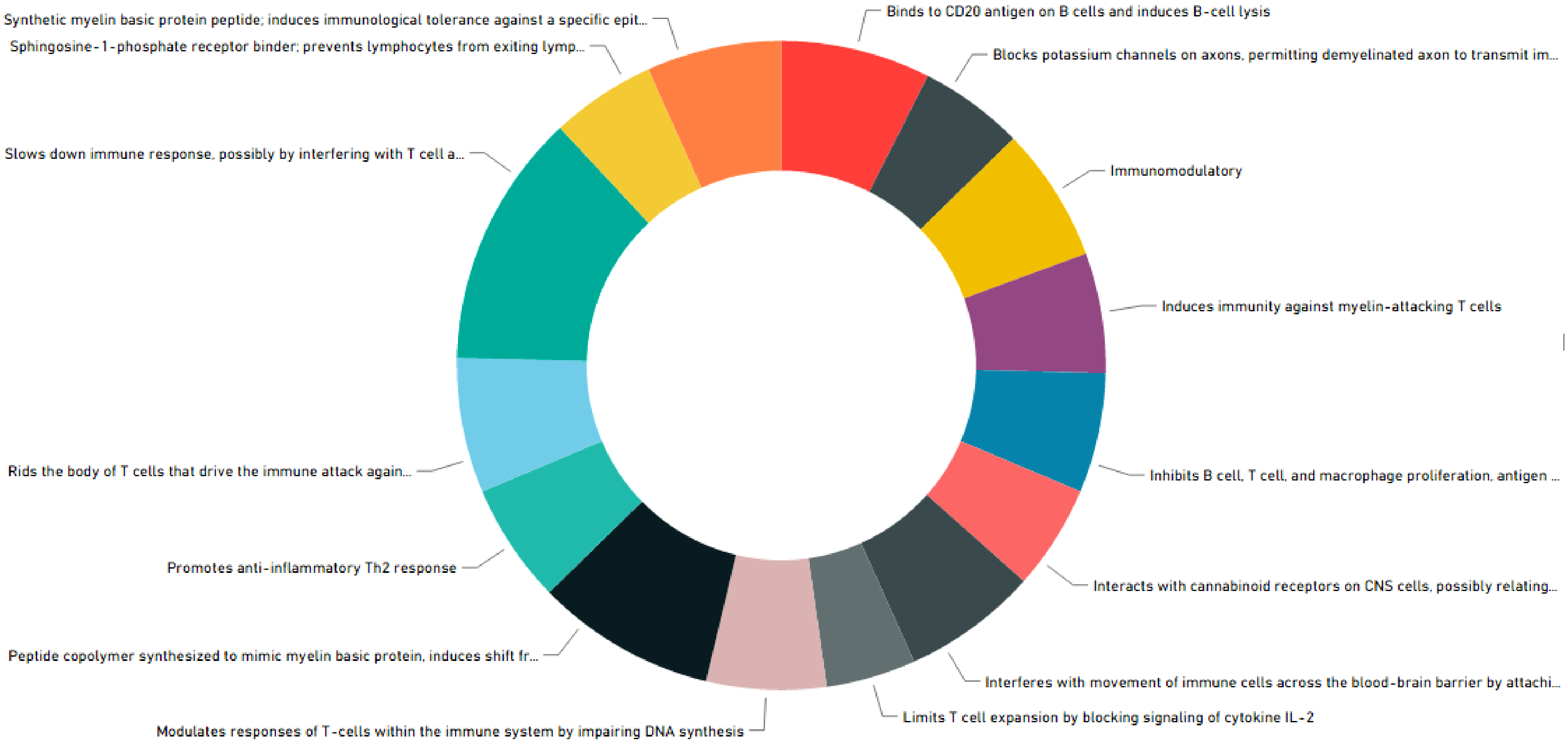


Figure 5.14: 15 top ranked mode of action of MS approved drugs, after processing clinical data taken from NMSS [244].

After developing the molecular interaction map of MS, the next task was to discover the specific pathways behind each of the disease subtypes in chronological order. The segregated pathways may involve to specific disease patterns thus we hypothesized that drugs targeting certain pathways may be combined to treat patients who are non-responders to some therapeutic agents. Combination therapy in MS is not a new topic, several papers [280–282] have been discussed this option to cure the disease but unfortunately the scope appears limited. Few studies were also carried out to find the efficacy of combinatorial therapies and one of them showed that IFN-Beta 1a and Glatiramer Acetate combination is not efficacious than IFN-Beta 1a alone [283]. In the following section, the pathways involved in different disease subtypes and the best possible combinatorial therapies with the help of pathways involved in different disease subtypes will be discussed.

5.2.2 Time series of MS disease progression and combinatorial therapy:

In this section our approaches to find out key elements behind each of MS disease subtype in time dependent manner have been described. In addition to finding disease mechanisms specific to a disease subtype, the possibility to use combinatorial therapy as a solution for non-responders will be explored by targeting pathways specific to the disease subtypes. To work with the stages of MS disease progression, the disease was segregated based on the biomarkers found in specific disease subtype and also the interacting molecules of the therapeutic agents. Following two different approaches have been used and outputs of both were compared to reveal the pathways which were commonly found or specific to a certain disease subtype:

- Literature based discovery
- Drug network based discovery

We speculated that if the outputs of both approaches are identical then it would confirm that the drugs were interacting with the same molecules perturbed in the disease subtype. On the contrary, if they are different, then a combination therapy can be proposed.

5.2.2.1 Literature based discovery:

In the literature-based approach, we collected the documented biomarkers associated with each stage of the disease by the methods described in previous chapter. For clarity, the workflow of the approach has been given:

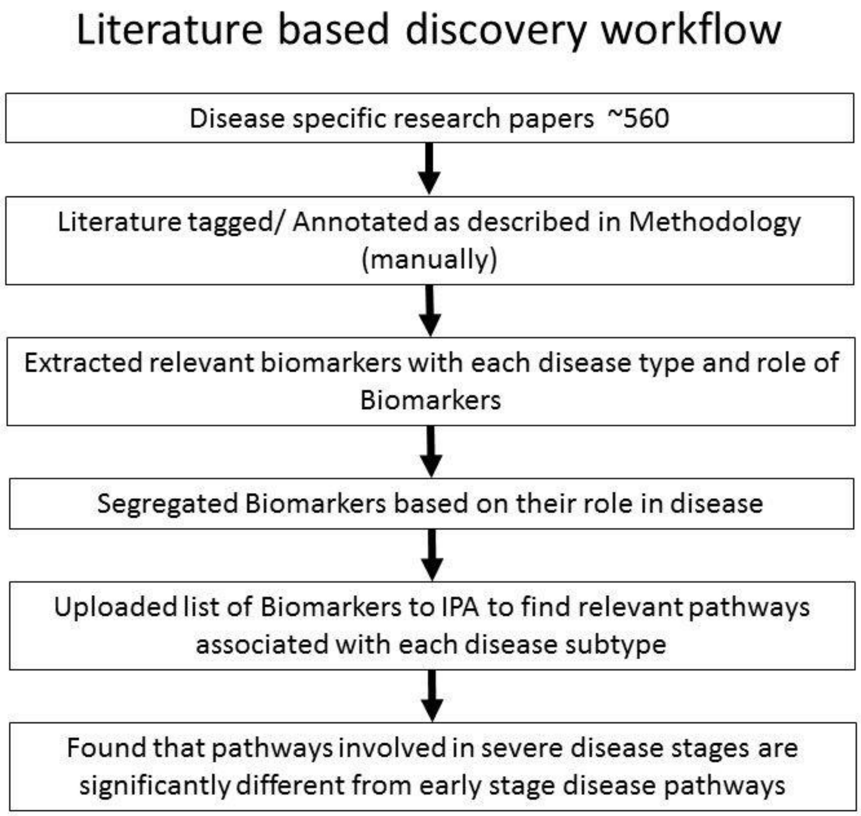

Literature based discovery workflow
Disease specific research papers ~560
Literature tagged/ Annotated as described in Methodology (manually)
Extracted relevant biomarkers with each disease type and role of Biomarkers
Segregated Biomarkers based on their role in disease
Uploaded list of Biomarkers to IPA to find relevant pathways associated with each disease subtype
Found that pathways involved in severe disease stages are significantly different from early stage disease pathways

Figure 5.15: Overall workflow of the approach for the literature-based discovery.

Figure 5.16 shows the relevant biomarkers mentioned with each disease subtype, mapped as their association was described in the literature.

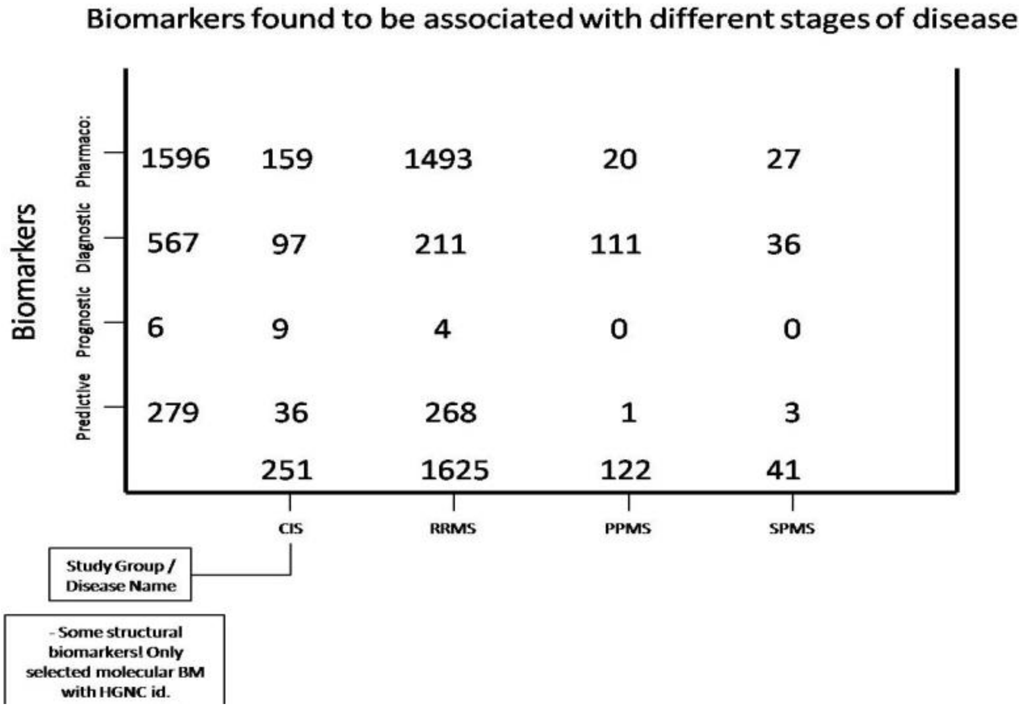
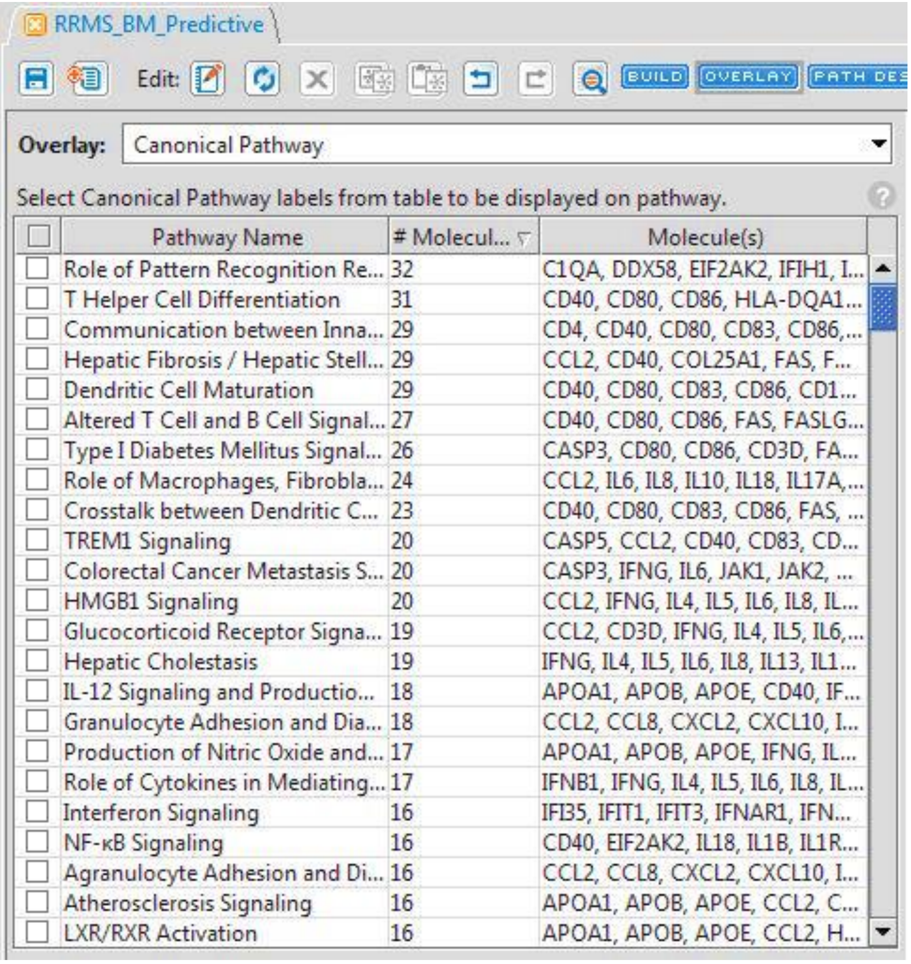


Figure 5.16: Different biomarkers grouped by type and associated with different disease subtypes. The total number of biomarkers is not the sum of rows, as some biomarkers were played more than one role in the disease.

There were many structural biomarkers specific for lesion types and their location, but we only considered molecular biomarkers with an HGNC identifier. We did not combine all the different type of biomarkers (Predictive, Prognostive and Diagnostive) as that would have been too many molecules and pathways. In addition, Pharmacovigilance biomarkers have not been used further as the main interest of the work was to discover disease mechanisms.

Only one biomarker was found to have a predictive role in the disease subtype PPMS, and only three in the disease subtype SPMS. This could be a reason why it is difficult to treat, slow down or even predict disease progression in advanced stages as well as not having a drug for PPMS so far.

Figure 5.17 shows the pathways mapped with predictive type biomarkers in disease subtype RRMS. Using the same method, all the relevant pathways have been mapped to all the biomarkers of disease subtypes e.g. CIS, RRMS (Predictive, Prognostive, Diagnostive), PPMS (Diagnostive), SPMS (Predictive, Diagnostive).



Overlay: Canonical Pathway

Select Canonical Pathway labels from table to be displayed on pathway.

□	Pathway Name	# Molecul...	Molecule(s)
□	Role of Pattern Recognition Re...	32	C1QA, DDX58, EIF2AK2, IFIH1, I...
□	T Helper Cell Differentiation	31	CD40, CD80, CD86, HLA-DQA1...
□	Communication between Inna...	29	CD4, CD40, CD80, CD83, CD86,...
□	Hepatic Fibrosis / Hepatic Stell...	29	CCL2, CD40, COL25A1, FAS, F...
□	Dendritic Cell Maturation	29	CD40, CD80, CD83, CD86, CD1...
□	Altered T Cell and B Cell Signal...	27	CD40, CD80, CD86, FAS, FASLG...
□	Type I Diabetes Mellitus Signal...	26	CASP3, CD80, CD86, CD3D, FA...
□	Role of Macrophages, Fibrobla...	24	CCL2, IL6, IL8, IL10, IL18, IL17A,...
□	Crosstalk between Dendritic C...	23	CD40, CD80, CD83, CD86, FAS, ...
□	TREM1 Signaling	20	CASP5, CCL2, CD40, CD83, CD...
□	Colorectal Cancer Metastasis S...	20	CASP3, IFNG, IL6, JAK1, JAK2, ...
□	HMGB1 Signaling	20	CCL2, IFNG, IL4, IL5, IL6, IL8, IL...
□	Glucocorticoid Receptor Signa...	19	CCL2, CD3D, IFNG, IL4, IL5, IL6,...
□	Hepatic Cholestasis	19	IFNG, IL4, IL5, IL6, IL8, IL13, IL1...
□	IL-12 Signaling and Productio...	18	APOA1, APOB, APOE, CD40, IF...
□	Granulocyte Adhesion and Dia...	18	CCL2, CCL8, CXCL2, CXCL10, I...
□	Production of Nitric Oxide and...	17	APOA1, APOB, APOE, IFNG, IL...
□	Role of Cytokines in Mediating...	17	IFNB1, IFNG, IL4, IL5, IL6, IL8, IL...
□	Interferon Signaling	16	IFI35, IFIT1, IFIT3, IFNAR1, IFN...
□	NF-κB Signaling	16	CD40, EIF2AK2, IL18, IL1B, IL1R...
□	Agranulocyte Adhesion and Di...	16	CCL2, CCL8, CXCL2, CXCL10, I...
□	Atherosclerosis Signaling	16	APOA1, APOB, APOE, CCL2, C...
□	LXR/RXR Activation	16	APOA1, APOB, APOE, CCL2, H...

Figure 5.17: Pathways retrieved after uploading list of RRMS Predictive Biomarkers into IPA.

5.2.2.2 Drug network based discovery:

Using a drug network-based approach, we aimed to discover chronological events occurrence in MS disease based on drugs prescribed. Following FDA-approved MS drugs have been selected and their molecular networks were constructed. The networks were mapped to pathways and top pathways were filtered out (Figure 5.18). The selection of drugs was made based on disease's state for which they are most commonly prescribed.

1. Interferon beta 1a
2. Interferon beta 1b
3. Glatiramer Acetate
4. Fingolimod
5. Natalizumab

6. Mitoxantrone
7. Methylprednisolone
8. Dimethyl Fumarate
9. Teriflunomide

The workflow of the approach to construct pathways linked with drugs is as follows:

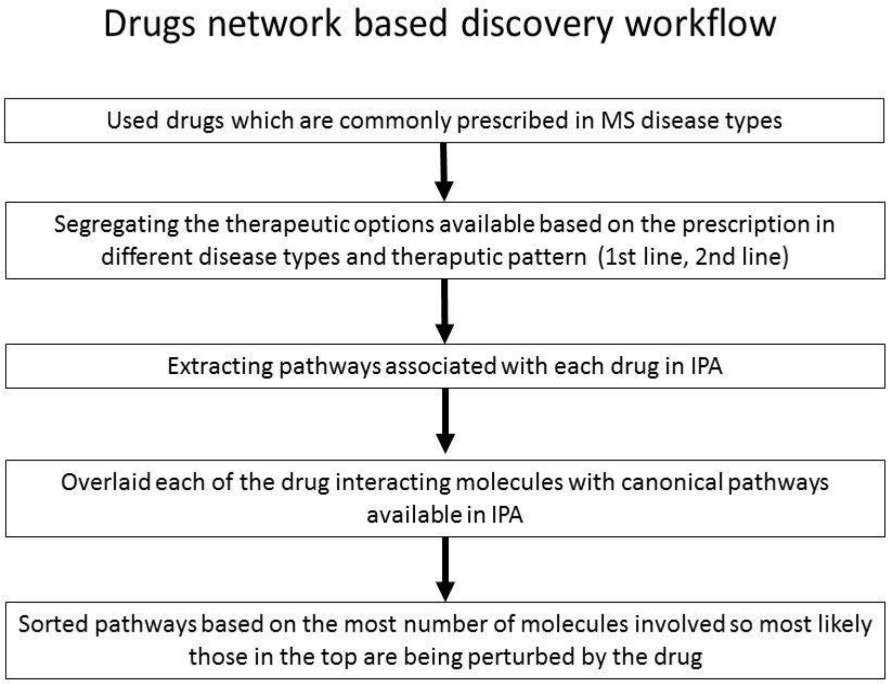

Drugs network based discovery workflow
Used drugs which are commonly prescribed in MS disease types
Segregating the therapeutic options available based on the prescription in different disease types and theraputic pattern (1st line, 2nd line)
Extracting pathways associated with each drug in IPA
Overlaid each of the drug interacting molecules with canonical pathways available in IPA
Sorted pathways based on the most number of molecules involved so most likely those in the top are being perturbed by the drug

Figure 5.18: Overall workflow of the method for drug network based discovery.

The drugs were segregated based on prescription pattern and therapeutic course e.g.1st line therapy and 2nd line therapy, as shown in figure 5.19. The 3rd line therapy is considered as combination therapy, while the 4th line therapy and final treatment option is a bone marrow transplant. The idea of segregation is to see if the drugs are perturbing the pathways intended.

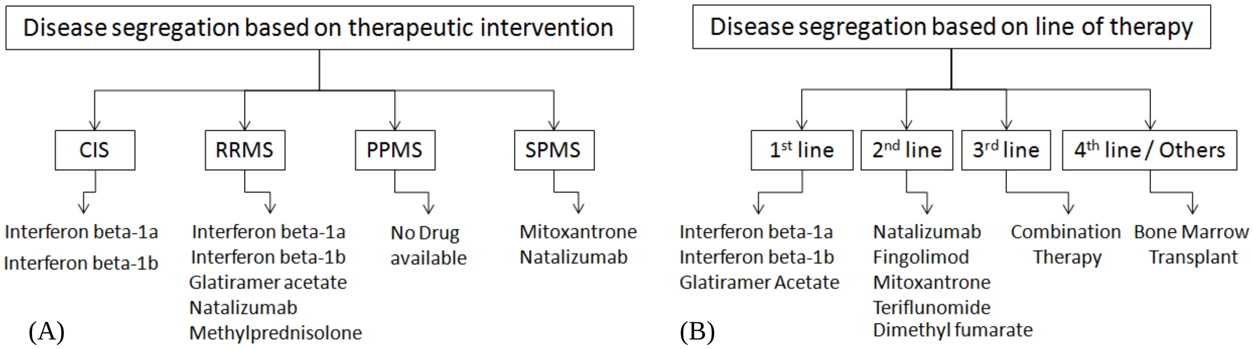
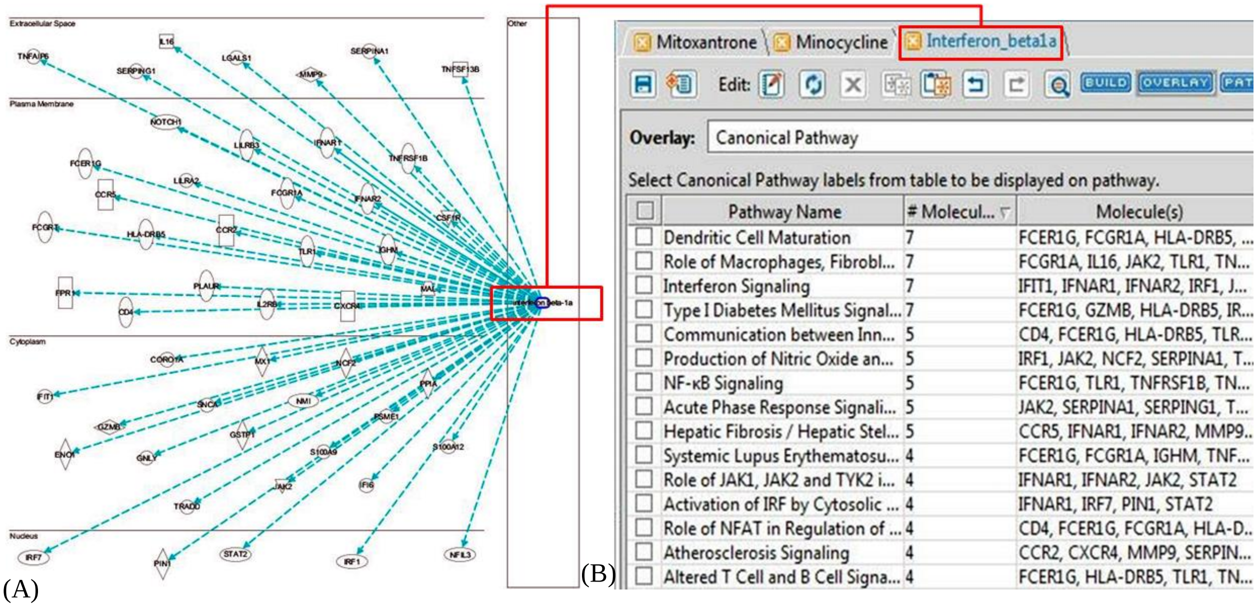


Figure 5.19: Two different disease segregation approaches based on therapies given at different stages and therapies considered as in chronological order were applied.

The drugs' networks were constructed with the knowledge available in IPA and then canonical pathways were mapped on the networks. The pathways were selected based on the maximum matches of molecules within the drug network; that is, the topmost pathway would be the one where the maximum number of molecules played key role in it. So theoretically, one could get all the pathways involved in a certain disease subtype by constructing the network of drugs prescribed (to treat or suppress the disease condition) and overlaying it with the canonical pathways. Figure 5.20 shows the details of the approach.



Overlay: Canonical Pathway

Select Canonical Pathway labels from table to be displayed on pathway.

□	Pathway Name	# Molecul...	Molecule(s)
□	Dendritic Cell Maturation	7	FCER1G, FCGR1A, HLA-DRB5, ...
□	Role of Macrophages, Fibrobl...	7	FCGR1A, IL16, JAK2, TLR1, TN...
□	Interferon Signaling	7	IFIT1, IFNAR1, IFNAR2, IRF1, J...
□	Type I Diabetes Mellitus Signal...	7	FCER1G, GZMB, HLA-DRB5, IR...
□	Communication between Inn...	5	CD4, FCER1G, HLA-DRB5, TLR...
□	Production of Nitric Oxide an...	5	IRF1, JAK2, NCF2, SERPINA1, T...
□	NF-κB Signaling	5	FCER1G, TLR1, TNFRSF1B, TN...
□	Acute Phase Response Signali...	5	JAK2, SERPINA1, SERPING1, T...
□	Hepatic Fibrosis / Hepatic Stel...	5	CCR5, IFNAR1, IFNAR2, MMP9..
□	Systemic Lupus Erythematosu...	4	FCER1G, FCGR1A, IGHM, TNF...
□	Role of JAK1, JAK2 and TYK2 i...	4	IFNAR1, IFNAR2, JAK2, STAT2
□	Activation of IRF by Cytosolic ...	4	IFNAR1, IRF7, PIN1, STAT2
□	Role of NFAT in Regulation of ...	4	CD4, FCER1G, FCGR1A, HLA-D..
□	Atherosclerosis Signaling	4	CCR2, CXCR4, MMP9, SERPIN...
□	Altered T Cell and B Cell Signa...	4	FCER1G, HLA-DRB5, TLR1, TN...

Figure 5.20: Drug network and pathway overlay: (A) Drug network retrieved from IPA knowledge base (B) Top associated canoncial pathways mapped with drug network.

Now, we have disease specific pathways (for each disease subtype) constructed from biomarkers and pathways constructed from drug networks. As the next step, a speculation was made to see whether or not there is a match between the two streams of knowledge. To test the speculation, comparisons of discovered pathways were performed. Figure 5.21 shows the simplified version of our comparison approach as we wanted to discover how many pathway matches there are, if any, and what role they play in different disease states.

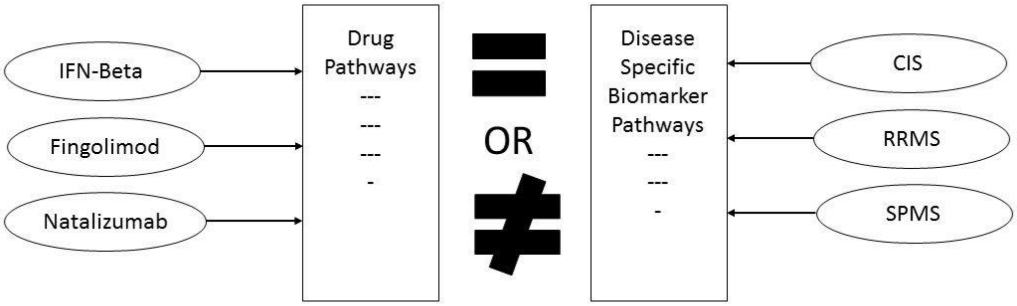


Figure 5.21: Summary of the question we were interested to look into.

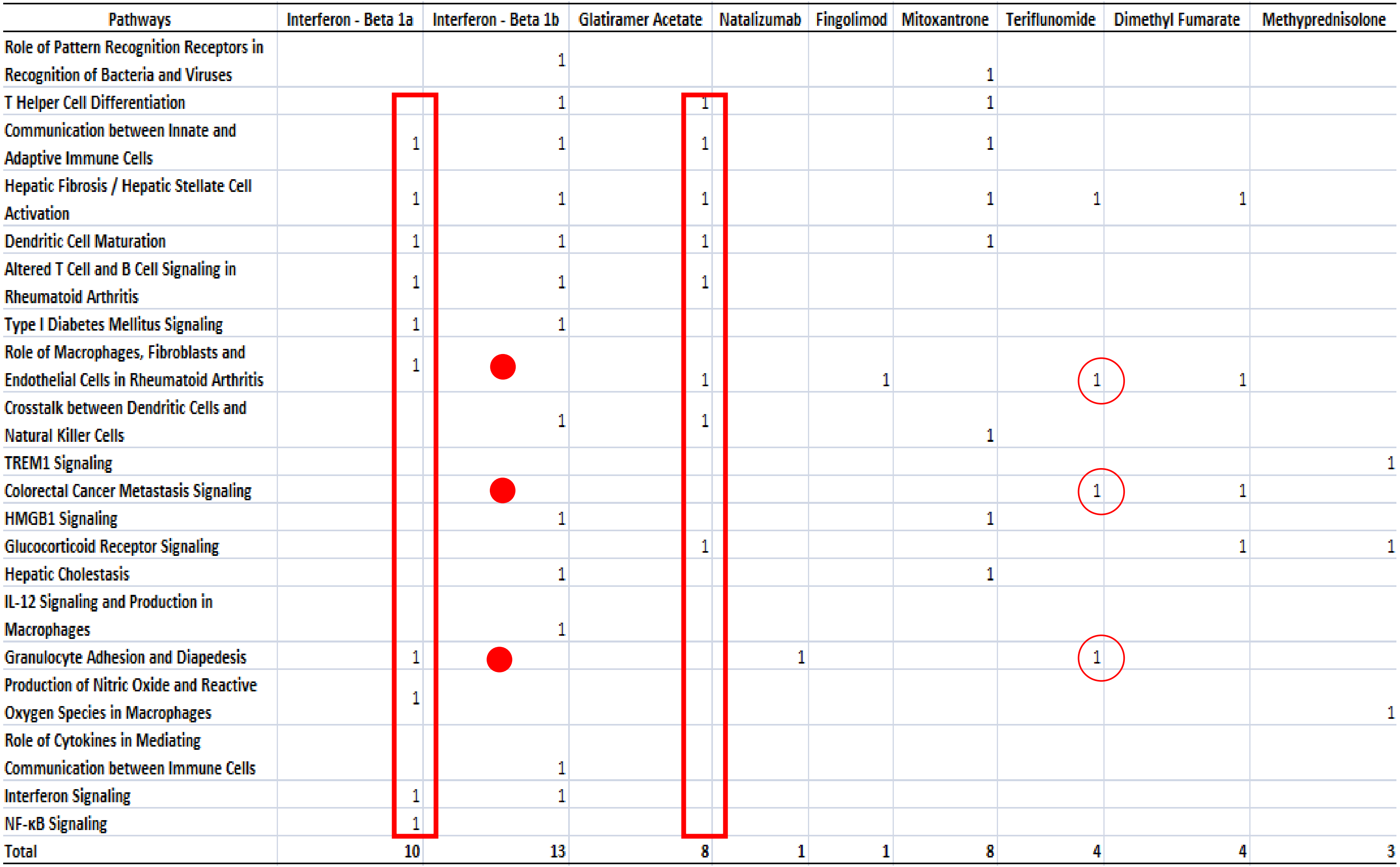
Top 20 associated pathways of drugs and disease subtypes have been chosen to find out which drugs have an interaction role in any of the disease pathways. The early stage disease pathways (associated with CIS & RRMS) were mostly perturbed by Interferon beta 1a and 1b, while Teriflunomide and Mitoxantrone were more frequently associated with the severe disease subtypes e.g. SPMS. In accordance with the established knowledge, it has been confirmed the chronological pattern of disease segregation based on therapeutic intervention. In addition to that it has also been found that combinatorial therapy could play a significant role to treat patients who are non-responders to certain drugs. The severe disease subtype pathways can be targeted by combining more than one drug. The approach is novel and unique as this is the first time to the author's knowledge that two established knowledge sets were used to discover the pathways responsible for a disease state.

Table 5.3 shows the number of pathways perturbed by different drugs and in different disease subtypes.

	IFN - Beta 1a	IFN - Beta 1b	Glatiramer Acetate	Natalizumab	Fingolimod	Mitoxantrone	Teriflunomide	Dimethyl Fumarate	Methyl-prednisolone
CIS	8	7	5	1	4	8	5	7	4
RRMS_ Predictive	10	13	8	1	1	8	4	4	3
RRMS_ Prognostive	7	5	4	1	1	5	1	5	3
RRMS_ Diagnostive	9	11	9	2	1	11	7	4	1
PPMS_ Diagnostive	9	9	8	0	1	9	5	5	3
SPMS_ Predictive	6	4	5	3	2	6	7	3	2
SPMS_ Diagnostive	7	6	10	3	2	8	9	5	4

Table 5.3: Pathways involved in different disease subtypes and drugs acting on those.

As discussed above that a study who looked at the combination therapy of IFN-Beta 1a and Glatiramer Acetate failed to prove that combination therapy is more efficacious than IFN-Beta 1a alone [283]. With the help of our work, we were able to explore that one of the possible reasons for unsuccessful study was that both of the drugs IFN-Beta 1a and Glatiramer Acetate interacted with similar pathways. 5 out of 8 pathways interacted by Glatiramer Acetate in disease subtype RRMS (mapped to biomarker type predictive) were also interacted by IFN-b1a. This shows that the drugs' mode of action was not non-overlapping (Figure 5.22). Figure 5.22 also shows that there are some pathways which are not interacted by some of the drugs and combining therapies based on those may yield beneficial outcome. However, experimental studies are needed to validate the findings.



Pathways	Interferon - Beta 1a	Interferon - Beta 1b	Glatiramer Acetate	Natalizumab	Fingolimod	Mitoxantrone	Teriflunomide	Dimethyl Fumarate	Methyprednisolone
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses		1				1			
T Helper Cell Differentiation		1	1			1			
Communication between Innate and Adaptive Immune Cells	1	1	1			1			
Hepatic Fibrosis / Hepatic Stellate Cell Activation	1	1	1			1	1	1	
Dendritic Cell Maturation	1	1	1			1			
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	1	1	1						
Type I Diabetes Mellitus Signaling	1	1							
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	1	●	1		1		1	1	
Crosstalk between Dendritic Cells and Natural Killer Cells		1	1			1			
TREM1 Signaling									1
Colorectal Cancer Metastasis Signaling		●					1	1	
HMGB1 Signaling		1				1			
Glucocorticoid Receptor Signaling			1					1	1
Hepatic Cholestasis		1				1			
IL-12 Signaling and Production in Macrophages		1							
Granulocyte Adhesion and Diapedesis	1	●		1			1		
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	1								1
Role of Cytokines in Mediating Communication between Immune Cells		1							
Interferon Signaling	1	1							
NF-κB Signaling	1								
Total	10	13	8	1	1	8	4	4	3

Figure 5.22: Pathways mapped to biomarkers found with RRMS Predictive type and the drugs interaction with the pathways. 5 out of 8 Pathways interacted by Glatiramer Acetate are same as IFN-b1a. The likelihood to have a effacacious combination therapy with those therapeutic agents is less than combining IFN-b1b and Teriflunomide.

Red filled circles show the not interacting pathways by IFN-b1b. 1 shows the pathways interacted by drugs. Red boxes show pathways of IFN-Beta 1a and Glatiramer Acetate.

Figure 5.23 shows the chart of pathways being interacted by drugs in different disease subtypes. Please note that disease subtype pathways were mapped according to their relevant biomarkers, the name corresponds to the same.

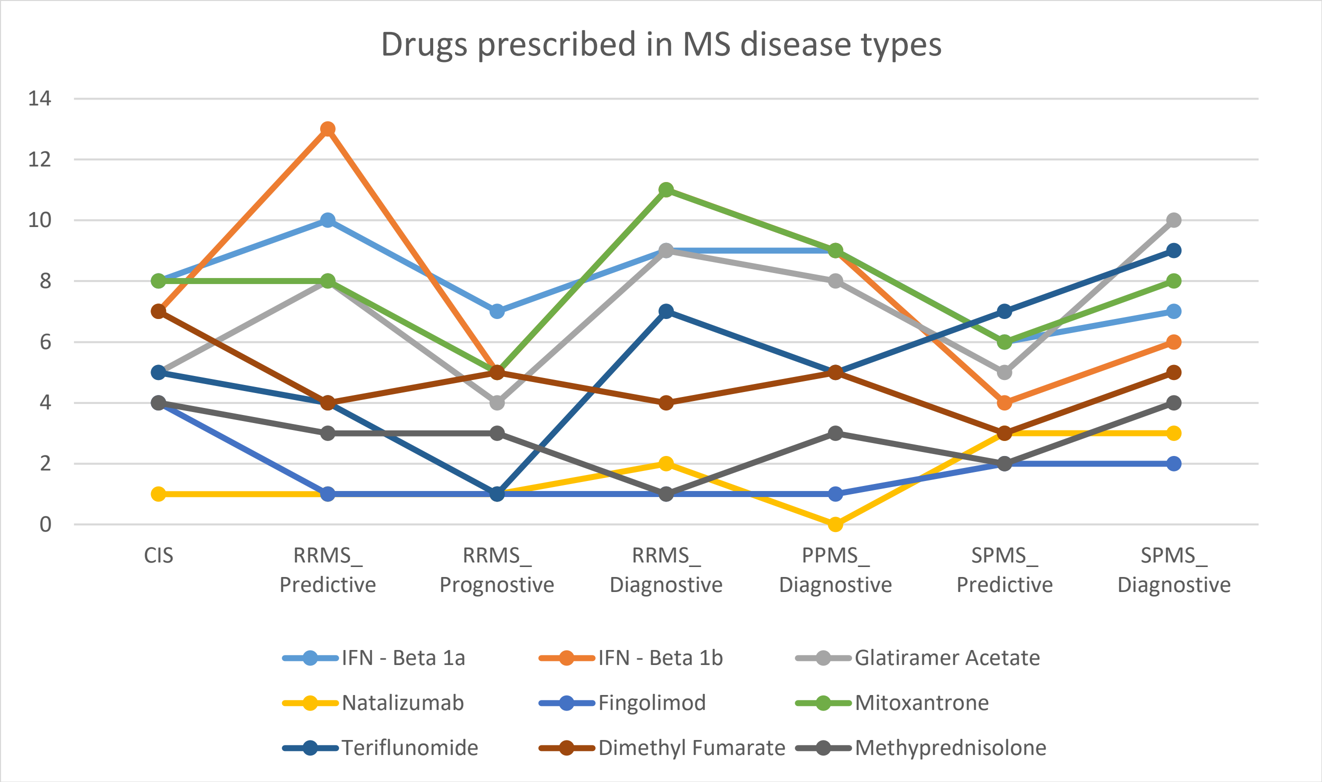


Figure 5.23: Disease subtype pathways being interacted by different drugs. Since the disease subtypes were segregated based on the associated biomarkers types, the pathways name corresponds to the same. IFN-Beta 1b interacts with the most number of pathways (with biomarker Predictive) in RRMS.

As a next step, all pathways of a disease subtype have been merged, resulting in pathways for RRMS instead of RRMS Predictive, RRMS Prognostive and RRMS Diagnostive. We assigned a value of one if a pathway played a role in only one stage of the disease (i.e. Diagnostive). If a pathway was found in all three roles of disease subtype RRMS, then it would receive a value of three. Only pathways which occurred more than once in any subtype of disease have been selected. Table 5.4 shows the pathways ranked according to their occurrences in the different disease subtypes. Some unique pathways of disease subtypes were also found e.g. IL-12 signaling pathway has a role only in the disease subtype RRMS; similarly, atherosclerosis signaling pathway has been shown in association with SPMS.

Pathways	RRMS	PPMS	SPMS
Hepatic Fibrosis / Hepatic Stellate Cell Activation	3		2
Glucocorticoid Receptor Signaling	3	1	2
Dendritic Cell Maturation	3	1	2
T Helper Cell Differentiation	3	1	
IL-12 Signaling and Production in Macrophages	2		
Crosstalk between Dendritic Cells and Natural Killer Cells	2	1	
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	2		
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	2	1	
HMGB1 Signaling	2		
Colorectal Cancer Metastasis Signaling	2	1	
Type I Diabetes Mellitus Signaling	2	1	
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	2	1	2
Role of Cytokines in Mediating Communication between Immune Cells	2		
Hepatic Cholestasis	2		
Communication between Innate and Adaptive Immune Cells	2	1	
Granulocyte Adhesion and Diapedesis	2		
IL-6 Signaling	2	2	
Leukocyte Extravasation Signaling			2
Agranulocyte Adhesion and Diapedesis			2
Atherosclerosis Signaling			2

Table 5.4: Pathways involved in different disease subtypes, only pathways which occurred more than once are shown here. Numbers here represent whether the same pathway is involved in all different types of biomarkers role (Predictive, Prognostive and Diagnostive).

In addition to the above mentioned pathways, we also looked at all pathways to reveal unique pathways specific to a disease subtype. Figures 5.24 and 5.25 show different pathways involved in different disease subtypes, and those which are unique for certain disease subtype. For this, only presence in certain disease subtype were observed and not the frequency. For example, if IL-6 signaling pathway was found twice in RRMS (with Predictive and Prognostive biomarker type) it was only considered that it has a role in RRMS thus given value of one to show that it plays a role in RRMS.

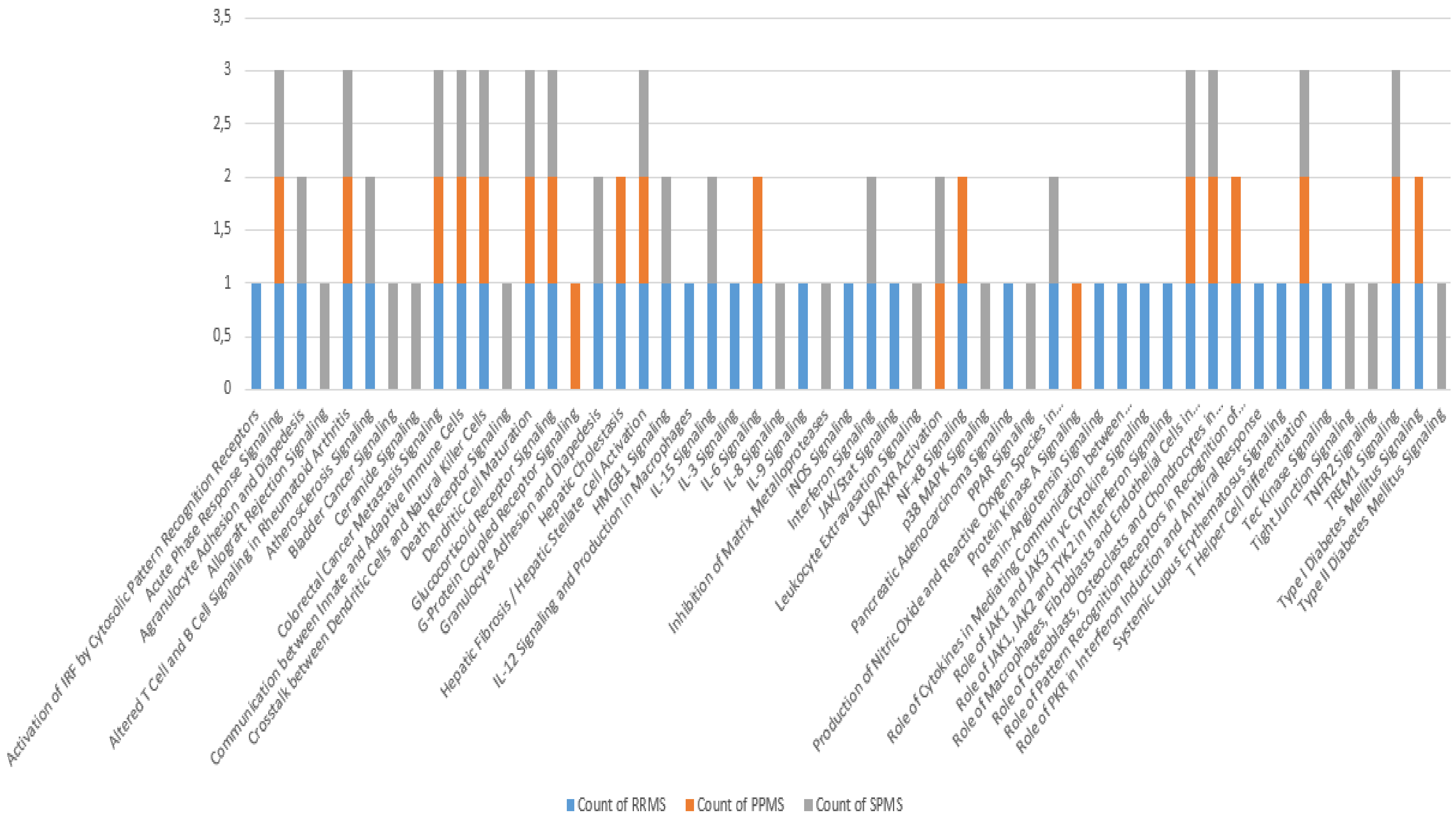


Figure 5.24: Pathways occurring in different disease subtypes. Only presence or absence was recorded to show whether or not a pathway plays a role in disease subtype.

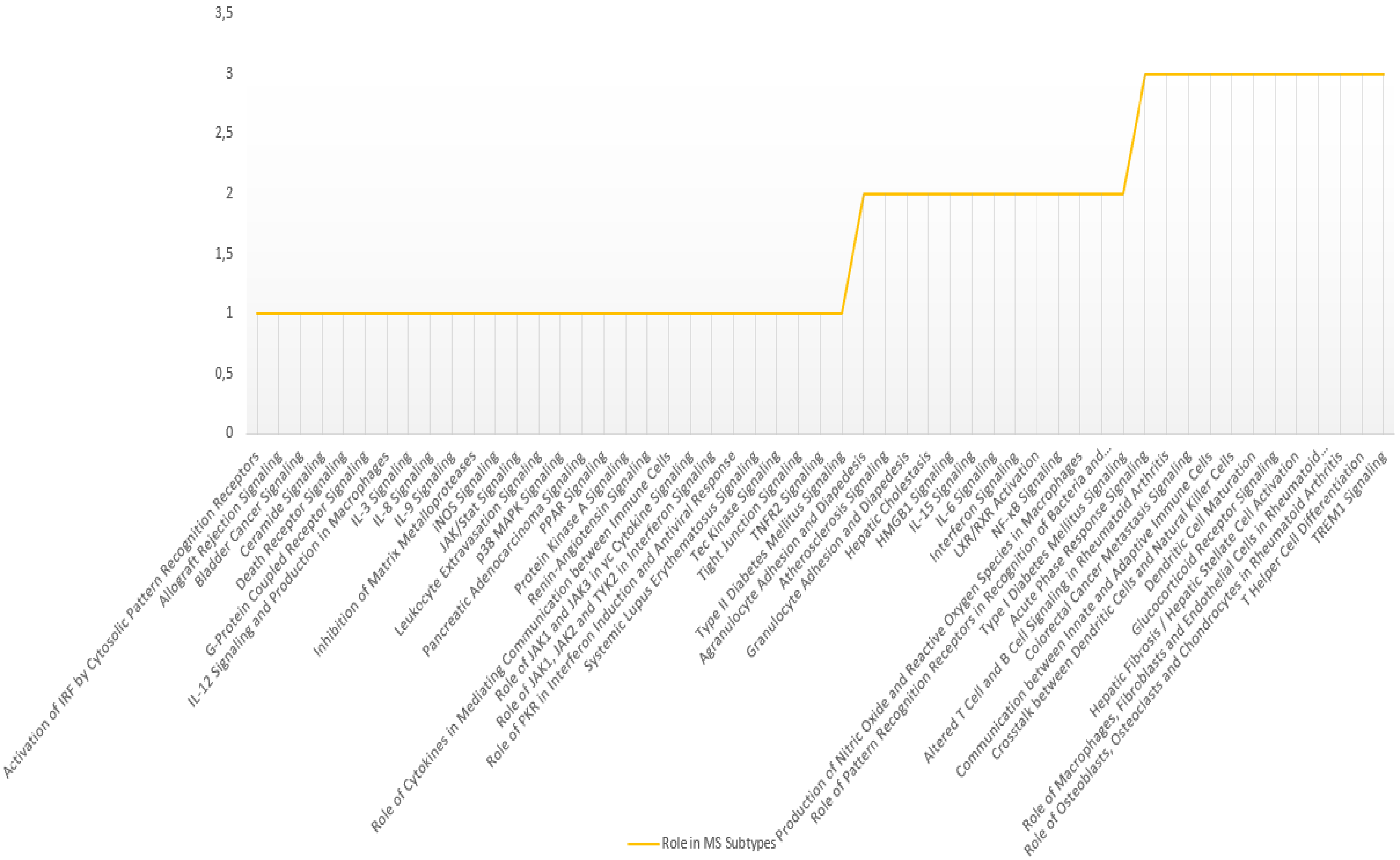


Figure 5.25: Frequency of pathways occurring in disease. Pathways which have a score of three means that they play a role in all the three disease subtypes i.e. RRMS, PPMS, and SPMS.

Both the most and least frequently occurring pathways have been depicted, in order to show the importance of certain pathways. By looking at the chart, one could see the most frequently occurring pathways in the less severe and in the more severe disease subtypes. In addition, patterns of pathways could also give a speculation about the disease progression and type in patients.

The next task was to develop a qualitative model of the disease where the relapsing remitting aspects of the disease can be simulated by using different factors involved. In next section, the result of modeling RRMS with Agent Based modeling is discussed.

5.2.3 Agent based model of Treg-Teff interplay and its role in RRMS: The Model and Simulation:

In this section, the results obtained from eight genetically predisposed, randomly chosen individuals have been presented. One aspect of this work is that it has been supposed that the appearance of a relapse and the presence of new unrecoverable neural damage are correlated [284], details are discussed in the previous chapter. The absence and presence of malfunctions of the Teff-Treg cross-balancing mechanisms at a local level have been reproduced. For simulating the absence of a local malfunction, it was supposed that both Teff and Treg populations had similar maximum duplication rates. In other words, we set the maximum duplication rate of Teff_dup and the duplication rate of Treg_pt to the same value, so both the cells’ populations have the same maximum duplication rates. It was further supposed that the breakdown of the cross regulation mechanism is due to a lower duplication rate pt of Treg. Table 5.5 shows the most important parameters used for the simulations.

Parameter	Meaning
treg_radius	max visibility radius of Treg
eff_dup	max. duplication rate of Teff
init_mye	initial quantity of myelin per patch
eat_mye	quantity of myelin destroyed by Teff
pt	max. duplication rate of Treg
patch_density	max. no of entities per patch allowed to have duplication
Teff_life	Teff mean half-life
Treg_life	Treg mean half-life

Table 5.5: Principal parameters of the MS agent based model.

The model has been tested by simulating 100 randomly chosen virtual patients (data not shown) in both the ill and the healthy scenarios by setting [eff_dup = 0.1; pt = 0.025] and [eff_dup = 0.1; pt = 0.1], respectively. The total damage has been noted for both of the scenarios at the end of the experiments. Median values of the final total damage were 77,268 for the ill sample and 5,357 for the healthy sample. Non-parametric Kolmogorov-Smirnov two samples goodness-of-fit test gave a maximum difference of D = 0.7500, (between the cumulative distributions) with a corresponding p-value of 0.000, thus suggesting that the two samples are unlikely to be drawn from the same distribution (i.e., they are statistically different).

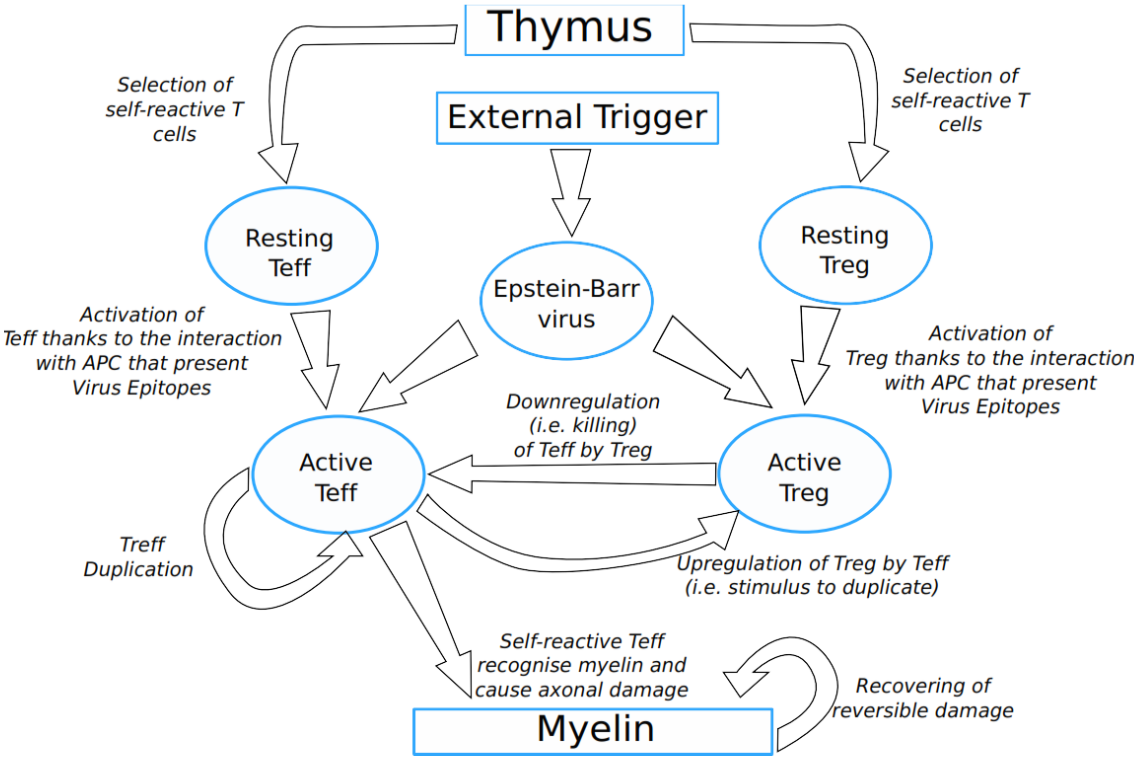

Thymus
Selection of self-reactive T cells
External Trigger
Selection of self-reactive T cells
Resting Teff
Epstein-Barr virus
Resting Treg
Activation of Teff thanks to the interaction with APC that present Virus Epitopes
Activation of Treg thanks to the interaction with APC that present Virus Epitopes
Downregulation (i.e. killing) of Teff by Treg
Active Teff
Active Treg
Treff Duplication
Upregulation of Treg by Teff (i.e. stimulus to duplicate)
Self-reactive Teff recognise myelin and cause axonal damage
Recovering of reversible damage
Myelin

Figure 5.26: The Conceptual Model of MS and the basis of the agent based model of MS.

Figure 5.27 shows the behavior of Teff, Treg, and viruses vs. time for all the presented individuals in absence of malfunctions of the Teff-Treg cross-balancing mechanisms. The number of Teff (in red) has spikes at different time intervals in all plots. This indicates that in some cases, due to the stochasticity in the introduction of newborn cells, self-reactive Teff may initially escape from Treg control (blue lines) and can be activated due to mimicry, duplicate, and try to attack myelin. However, activated Treg are able to counterbalance Teff actions and maintain immune homeostasis.

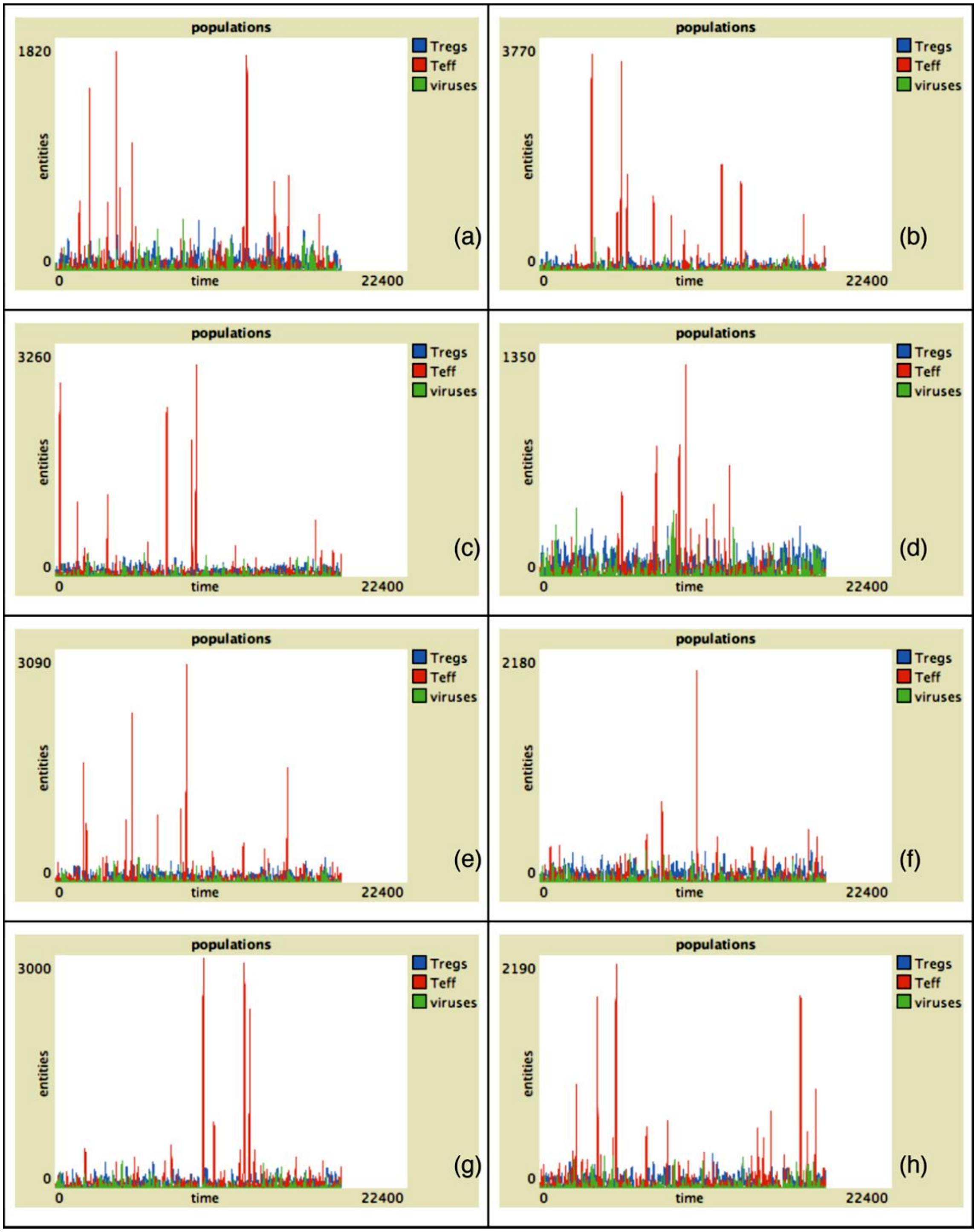


Figure 5.27: Entity behaviors vs. time in healthy patients. The simulation is based on eight randomly-selected healthy virtual patients. Similar duplication rates have been assumed for both Treg and Teff. Simulation time is 5 years (18,250 time-steps). Red lines represent activated Teff behaviors, blue lines represent activated Treg behaviors and green lines represent viruses' behaviors. In this case, the number of Teff peaks is relatively small due to the action of regulatory mechanisms. This would result in lower probabilities of having unrecoverable damage.

In Figure 5.28, the behavior of Teff, Treg and viruses vs. time for all the individuals in the case of malfunctions of the Teff-Treg cross-balancing mechanisms has been shown. Similar to what has been observed in Figure 5.27, all plots show some spikes in the Teff behaviors (red lines). However, in this case, the spikes are more numerous and reach higher values than expected. This suggests that, due to the malfunction in the regulatory mechanisms, Teff can be easily activated and cause brain damage. In this case, Treg are not always able to contrast Teff actions and maintain homeostasis.

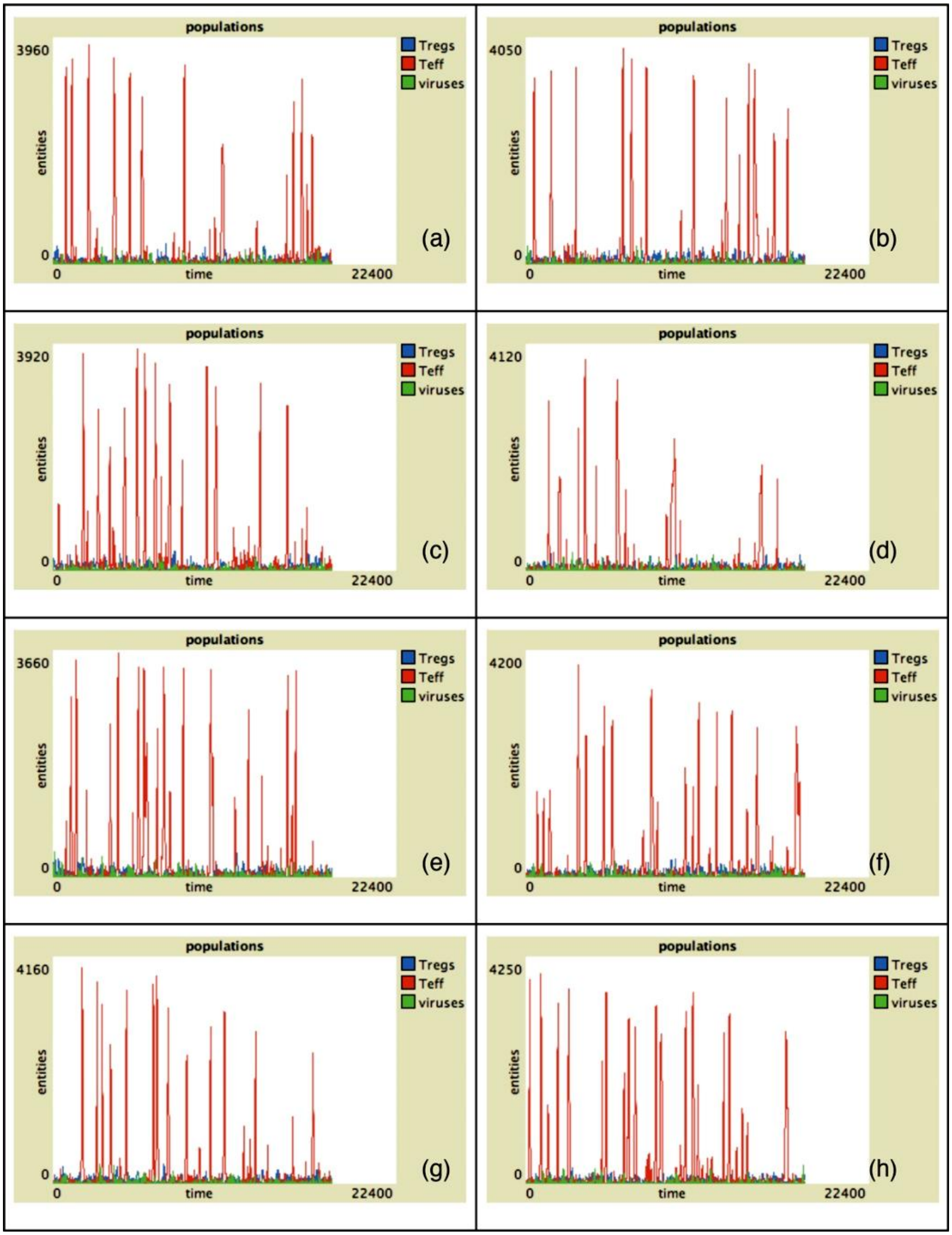


Figure 5.28: Entity behaviors vs. time in ill patients. The simulation is based on eight randomly-selected ill virtual patients. It was supposed that the breakdown of the cross regulation mechanism is due to a lower duplication rate pt of Treg. Simulation time is 5 years (18250 time-steps). Red lines represent activated Teff behaviors, blue lines represent activated Treg behaviors and green lines represent viruses' behaviors. In this case, the number of Teff peaks is higher. Moreover, each peak reaches higher values with respect to healthy patients, thus indicating that higher numbers of self-reactive Teff may entitle higher probabilities of having unrecoverable damage.

In figure 5.29, the levels of damage (recoverable, unrecoverable and total) for all the simulations have been shown. The presence of some spikes in the recoverable damage plots (blue lines) indicates that these correspond to the Teff spikes shown in figure 5.27. However, such damage is usually recovered, and at the end of this simulation almost no unrecoverable damage (red lines) remains.

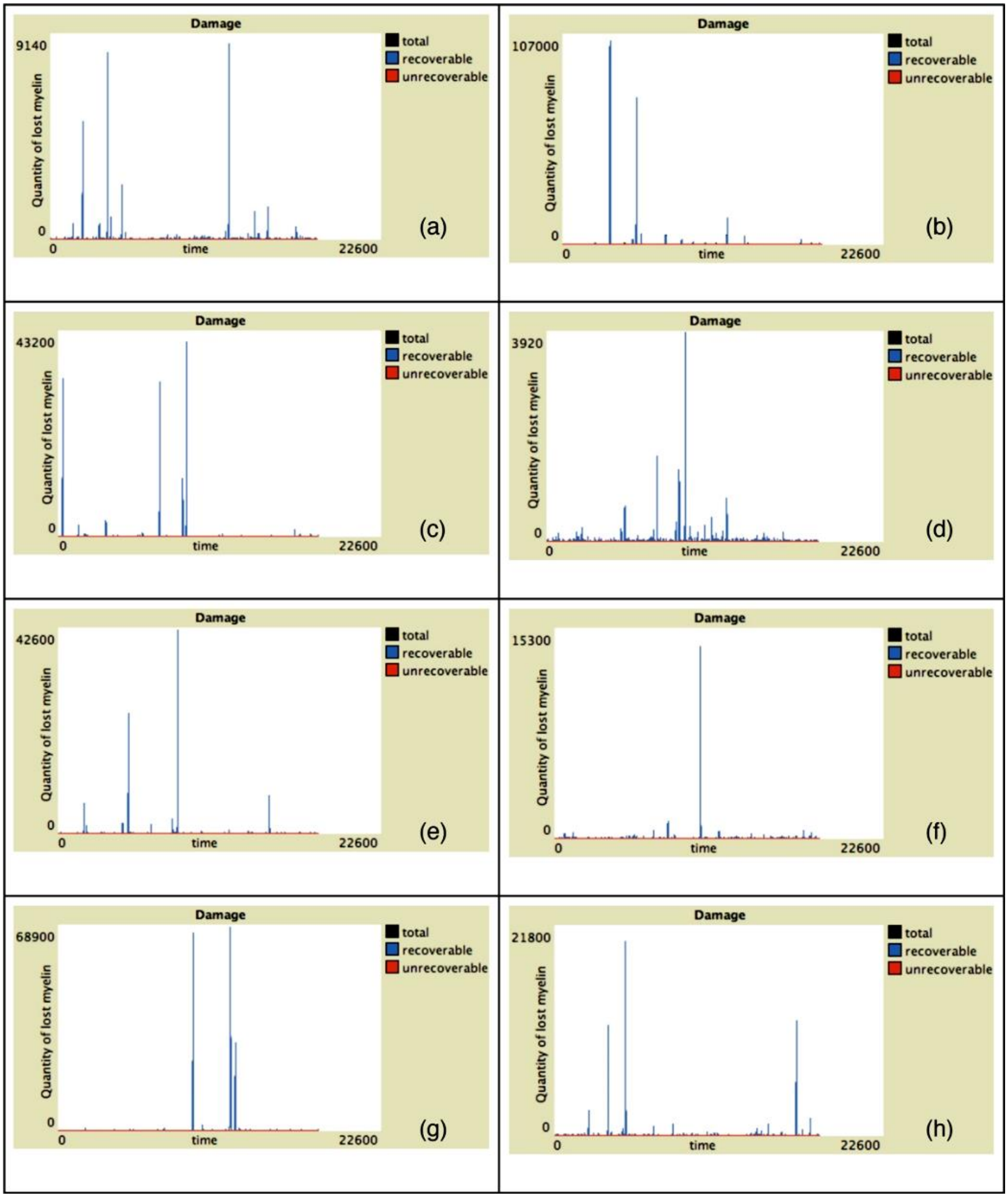


Figure 5.29: Damage progression vs. time in healthy patients. The simulation is based on eight randomly-selected healthy virtual patients. Simulation time is 5 years (18250 time-steps). Red lines represent activated unrecoverable damage, blue lines represent recoverable damage and black lines represent total damage (recoverable + unrecoverable). Some spikes on the recoverable damage curves are present. However, such damage is usually recovered in healthy patients, as at the end of simulations the total damage is mostly zero.

In figure 5.30, the levels of damage are shown for all the simulations. The spikes in the recoverable damage plots (blue lines) are higher and bigger in number. It is also possible to observe the appearance of unrecoverable damage that indicates the appearance of MS plaques, and to see how the sum of both (total damage, black plots) mimics the typical relapsing-remitting dynamics observed in MS.

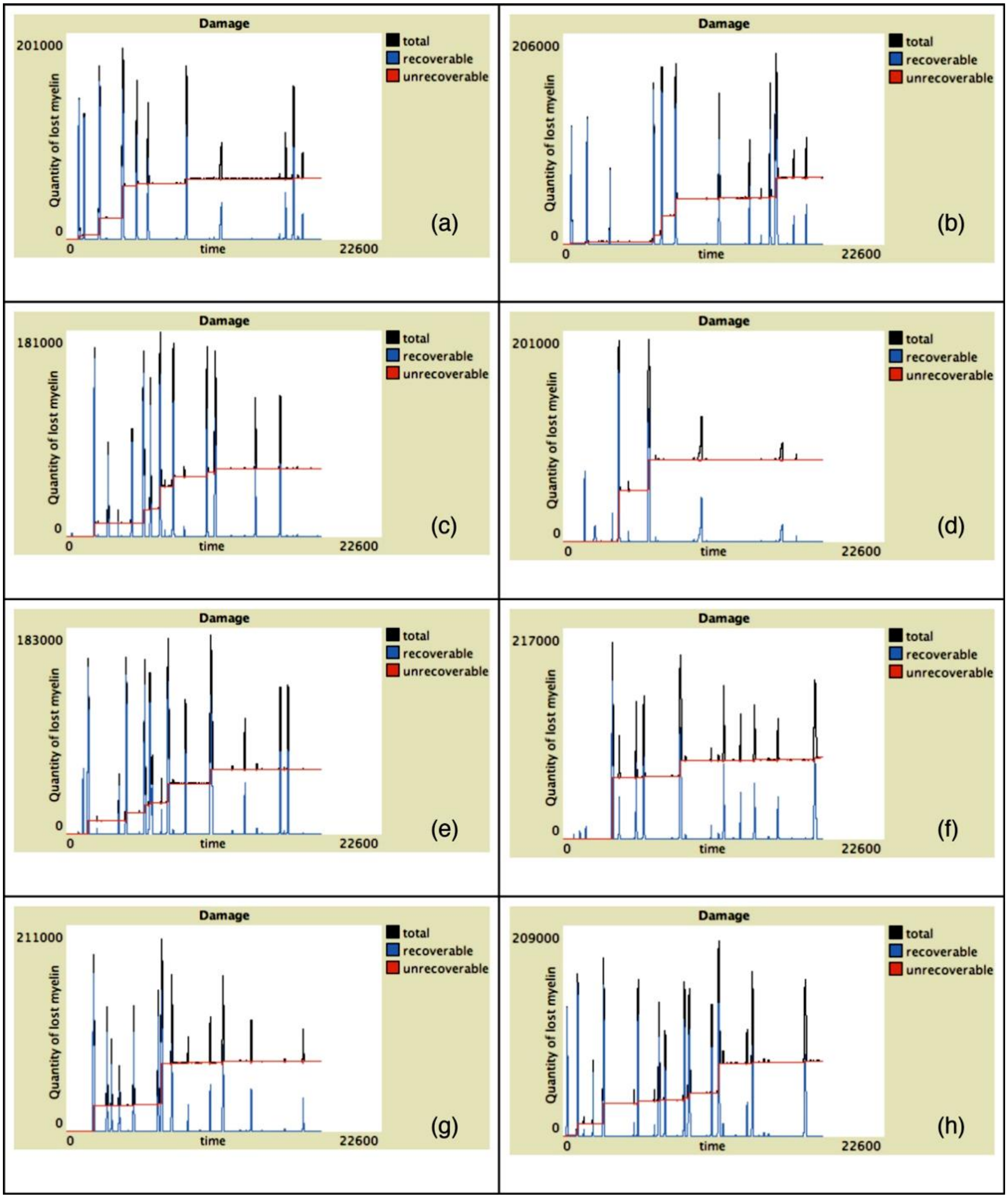


Figure 5.30: Damage progression vs. time in ill patients. The simulation is based on eight randomly-selected ill virtual patients. Simulation time is 5 years (18250 time-steps). Red lines represent activated unrecoverable damage, blue lines represent recoverable damage and black lines represent total damage (recoverable + unrecoverable). In this case it is possible to observe more frequent spikes in the recoverable damage curves. Furthermore, unrecoverable damage (that can be correlated with the appearing MS plaques) is also present.

Figure 5.31 presents the spatial plots at the end of every simulation (after 5 years) in healthy patients. The plots confirm the observations that came from Figure 5.27, as almost no black patches (which indicate the presence of some scarring or lesions) are present.

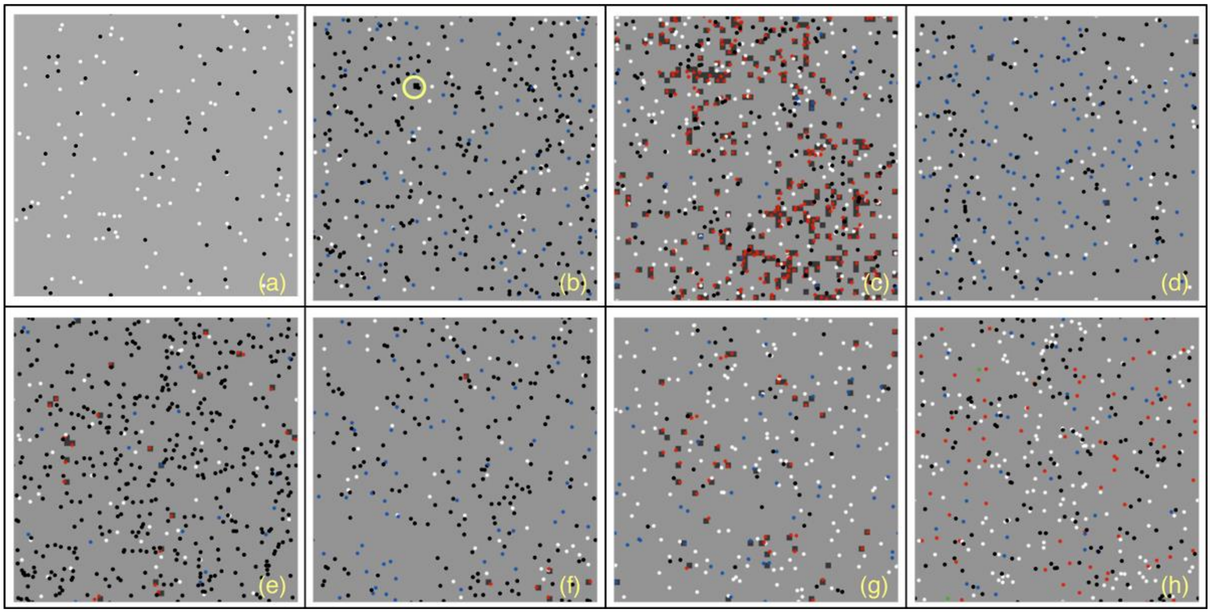


Figure 5.31: Spatial plot at the end of the simulation in healthy patients. The figure gives a spatial representation of the simulated scenario (i.e. a small portion of brain tissues) at the end of the simulation for eight randomly-selected healthy virtual patients. Light green patches represent non-damaged areas. Dark green patches (see for example plot (c)) represent areas with recoverable damage. Black patches (see for example plot (b), yellow circle) represent areas with unrecoverable damage. Red dots represent activated Teff and blue dots represent activated Treg. Green dots represent viruses. White and black dots represent resting Teff and Treg, respectively.

In figure 5.32, the presence of scarring can be seen, where the spatial plots for all individuals are presented at the end of 5 years. In all plots it is possible to see many black areas that indicate unrecoverable damage and thus the presence of lesions and scarring that may be correlated with relapses and the appearance of disability.

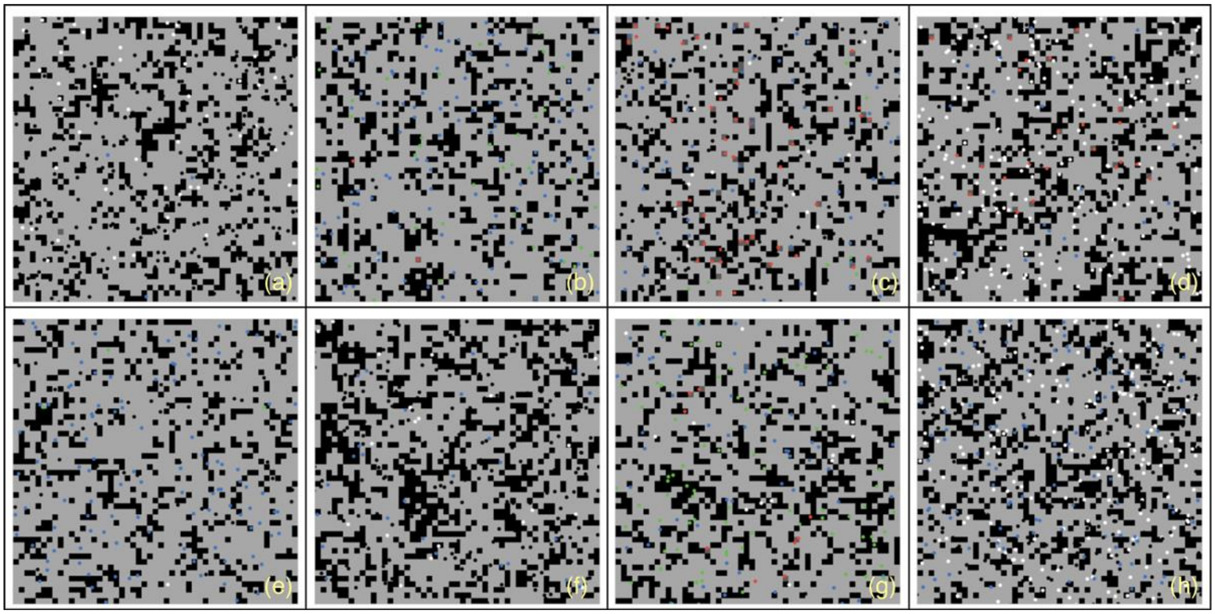


Figure 5.32: Spatial plot at the end of the simulation in ill patients. Figure gives a spatial representation of the simulated scenario (i.e. a small portion of brain tissues) at the end of the simulation for eight randomly-selected healthy virtual patients. Light green patches represent non-damaged areas. Dark green patches represent areas with recoverable damage. Black patches represent areas with unrecoverable damage. Red dots represent activated Teff and blue dots represent activated Treg. Green dots represent viruses. White and black dots represent resting Teff and Treg, respectively.

We also observed that (data not shown), in some cases (for some seeds) a decrease in the Teff or an increase in Treg proliferation does not always indicate less severe relapses and, as a matter of fact, it could produce more severe relapses. This is mainly due to the stochasticity of the model. It may happen that the stochastic injection of new resting Teff may not be shortly followed by an equivalent injection of Treg, which would create a temporary disequilibrium between the two populations; this would result in some neural damage even in potentially healthy patients.

The results presented here suggest that the presence of a genetic predisposition is not always a sufficient condition for developing the disease. Other conditions such as a breakdown of the mechanisms that regulate and allow peripheral tolerance should be involved. This has also been observed in [285]. In our case, we supposed that a malfunction of self-reactive regulatory T-cells caused by lower duplication rates was the cause. Of course, other conditions may be the cause of such a malfunction.

Moreover, it was also observed that in the simulations of ill patients, relapses mainly occurred in the first half of the simulation rather than in the second half (see Figure 5.30, plots (a),(c),(d) and (e)). This could be in line with clinical observations which showed that the relapse rate tends to decrease as the disease progresses [286,287].

6 Conclusion and Outlook:

Drug discovery with the help of Systems biology is a modern approach, as the conventional means of drug discovery are not efficient for finding treatments for polygenic diseases and there is a large unmet medical need. Polygenic diseases affect many systems of the body, thus studying the root cause as well as the chronic nature of the disease progression is crucial to finding a cure. The diseases of the human brain are among the most complicated polygenic diseases, which is unsurprising given that it is ranked as the most complicated object in the known universe. Unlike other organs, it is not directly accessible for observation e.g. ears, eyes, nose or skin, and it never stops functioning throughout life as it plays a crucial role in nervous system. Since the organ and its diseases are so complex, a systematic view of the diseases could be considered as an appropriate approach to find therapeutic treatments. In addition, the emergence of new technologies in biomedicine and the availability of vast variety of datasets with different formats make it challenging to have a clear picture of biological mechanisms, unless an integrated approach is considered. Systems biology provides the tools and methods to unravel the mechanisms behind complex phenomena.

Multiple sclerosis, being one of the most complex and expensive neurodegenerative diseases, has unmet medical needs due to its unknown etiology and varying pattern of progression. The disease affects mostly young people between 20-40 years of age and in an observable chronology renders patients disabled. To find the underlying disease mechanisms and biological phenomena, systems biological approaches have been used to discover the pathways involved in the disease and to seek a cure for the disease.

The modeling of diseases has been recently gained popularity after the successful acceptance of type 1 diabetes metabolic simulator (T1DMS) by the FDA as a substitute for pre-clinical animal testing of new treatment strategies for type 1 diabetes mellitus [154]. In-silico models allow the changing of parameters fairly easily in order to affect the variations in outcomes. We developed models to simulate the disease conditions in-silico and foster drug discovery by analyzing those models. These models were developed with different approaches and mainly based on manually curated knowledge excerpted from scientific literature.

The approach we used allowed us to systematically capture knowledge obscured in the literature and helped us to find unique patterns of disease foundation, its signatures and progress patterns. Besides disease modeling, a disease-specific ontology was developed to have a context-based search of concepts and phenotypes associated with them. Further, the enrichment of the concepts with relevant Spanish terms made it possible to extract valuable knowledge from the Spanish dataset. The methodology to enrich ontologies with the synonyms from 23 languages has been derived and used in our work. A software program was also developed to ease ontology development with essential concepts integration so that a framework of information to knowledge has been projected. Information can be gathered about any specific topic or any specific disease with the guidance of the methodology described in our paper [30]. The concept synonyms can be retrieved from one of the largest repository of medical language or any MySQL database as described in our paper [29]. The translation of the ontology is also possible, as UMLS provides support for many languages. The software tool helps to integrate concepts' synonyms and relevant information, and convert all into an ontology [208]. The framework not only develops de novo ontologies but can hypothetically enrich and translate any existing ontology. This answers our question of how to transform information into an integrated knowledge base. The following achievements of the thesis are given:

- A methodology to extract biomedical information from publically available data as well as proprietary data (e.g. Clinical Trials or UMLS database), and transform it into a structured ontology; this includes the possibility to enrich and/or translate it into any of 23 languages, thus broadening the ontology-based information extraction and coverage.
- Knowledge-based models of multiple sclerosis in the form of molecular interaction maps, disease ontology and an agent based dynamic model. The disease ontology provided significant correlation between multiple sclerosis, showing relations with other diseases and its role in comorbidities. The model provided disease simulation; it has been shown that an agent-based model could simulate virtual patient symptoms of relapsing remitting multiple sclerosis and the protective role of vitamin D. A time pattern of disease progression has also been modeled, and it has been shown that different biomarkers and unique pathways are involved in different stages of disease with respect to time. In addition, drugs' interactions were explored and molecules which interacted with more than one drug were investigated. The models developed can be helpful tools to find an appropriate drug candidate for the cure of the MS disease.
- A methodology and application to transform molecular interaction models or any SBML file into a Media wiki-based knowledge base. The application allows users to

connect a semantic knowledge base with internal and external databases and to enrich the models with information from any linked data platform.

We have discovered novel pathways about each of the different disease subtypes and a future course of action would be to design drugs based on the molecules and biomarkers involved in the specific disease subtypes. Further, these molecules' roles must be observed in human clinical trials. The current set of data is taken from published scientific literature where most of the findings were established on clinical animals; molecular interaction may vary in human subjects.

The obvious next step after the modeling would have been the validation with clinical and experimental data; both are present at Merck and collaboration partners of Merck in the MS area; however the responsible people in the indication area have not been prepared to share the data thus approaches could not be validating in living organisms. The approaches we used can be applied to any disease and they open up a new prospect of discovering drugs.

7 References:

1. Funahashi A, Matsuoka Y, Jouraku A, Morohashi M, Kikuchi N, Kitano H. CellDesigner 3.5: a versatile modeling tool for biochemical networks. Proc IEEE. 2008;96:1254–65.

2. Kitano H, Funahashi A, Matsuoka Y, Oda K. Using process diagrams for the graphical representation of biological networks. Nat. Biotechnol. 2005;23:961–6.

3. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34:D668–72.

4. Ingenuity IPA - Integrate and understand complex ’omics data [Internet]. Ingenuity. [cited 2016 Jun 1]. Available from: http://www.ingenuity.com/products/ipa

5. KNIME | Open for Innovation [Internet]. [cited 2014 Dec 11]. Available from: https://www.knime.org/

6. TEMIS launches Luxid® Content Enrichment Platform 7.1 - [Internet]. 2015 [cited 2017 Nov 3]. Available from: http://www.expertsystem.com/temis-launches-luxid-content-enrichment-platform-7-1/

7. MediaWiki [Internet]. [cited 2017 Nov 3]. Available from: https://www.mediawiki.org/wiki/MediaWiki

8. Le Novère N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, et al. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat. Biotechnol. 2005;23:1509–15.

9. MySQL [Internet]. [cited 2017 Nov 3]. Available from: https://www.mysql.com/

10. NetLogo Home Page [Internet]. [cited 2017 Nov 3]. Available from: https://ccl.northwestern.edu/netlogo/

11. Rajput A-M, Pennisi M, Motta S, Pappalardo F. OntoFast: Construct Ontology Rapidly. In: Klinov P, Mouromtsev D, editors. Knowl. Eng. Semantic Web [Internet]. Springer International Publishing; 2014 [cited 2016 Sep 19]. p. 237–41. Available from: http://link.springer.com/chapter/10.1007/978-3-319-11716-4_21

12. Musen MA. The Protégé Project: A Look Back and a Look Forward. AI Matters. 2015;1:4–12.

13. Rajput AM, s TM, Haase P, Scheer A, Toldo L. SBML2SMW Links Systems Biology, Text Mining and Semantic Web. In: Bichindaritz I, Perner P, s GR, Schmidt R, editors. IBaI Publishing; 2012. p. 177–84.

14. Fraunhofer SCAI: SCAIView [Internet]. 2013 [cited 2013 Feb 26]. Available from: http://www.scai.fraunhofer.de/en/business-research-areas/bioinformatics/products/scaiview.html

15. Unified Medical Language System (UMLS) [Internet]. [cited 2017 Nov 3]. Available from: https://www.nlm.nih.gov/research/umls/

16. Xu R, Musen MA, Shah NH. A comprehensive analysis of five million UMLS metathesaurus terms using eighteen million MEDLINE citations. AMIA. Annu. Symp. Proc. [Internet]. American Medical Informatics Association; 2010 [cited 2015 Feb 3]. p. 907. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041393/

17. Programming language [Internet]. Wikipedia. 2017 [cited 2017 Oct 23]. Available from: https://en.wikipedia.org/w/index.php?title=Programming_language&oldid=806395519

18. Azevedo FAC, Carvalho LRB, Grinberg LT, Farfel JM, Ferretti REL, Leite REP, et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J. Comp. Neurol. 2009;513:532–41.

19. Kosinski R, Kosinski RA, Zaremba M. Dynamics of the Model of the Caenorhabditis elegans Neural Network. Acta Phys. Pol. B. 2007;38:2201.

20. Yuste R, Church GM. The new century of the brain. Sci. Am. 2014;310:38–45.

21. D'Orazio J, Jarrett S, Amaro-Ortiz A, Scott T. UV Radiation and the Skin. Int. J. Mol. Sci. 2013;14:12222–48.

22. The Human Brain Project - Human Brain Project [Internet]. [cited 2014 Dec 30]. Available from: https://www.humanbrainproject.eu/

23. Brain Research through Advancing Innovative Neurotechnologies (BRAIN) - National Institutes of Health (NIH) [Internet]. [cited 2014 Dec 30]. Available from: http://braininitiative.nih.gov/

24. Allen Brain Atlas - Home [Internet]. [cited 2014 Dec 30]. Available from: http://www.brain-map.org/

25. Bluebrain | EPFL [Internet]. [cited 2014 Dec 30]. Available from: http://bluebrain.epfl.ch/

26. BRAINMAPS.ORG - BRAIN ATLAS, BRAIN MAPS, BRAIN STRUCTURE, NEUROINFORMATICS, BRAIN, STEREOTAXIC ATLAS, NEUROSCIENCE [Internet]. [cited 2014 Dec 30]. Available from: http://brainmaps.org/

27. Atlas of MS [Internet]. MS Int. Fed. [cited 2014 Dec 30]. Available from: http://www.msif.org/about-us/advocacy/atlas-of-ms/

28. Malhotra A, Gündel M, Rajput AM, Mevissen H-T, Saiz A, Pastor X, et al. Knowledge Retrieval from PubMed Abstracts and Electronic Medical Records with the Multiple Sclerosis Ontology. PLoS ONE. 2015;10:e0116718.

29. Rajput AM, Gurulingappa H. Semi-automatic Approach for Ontology Enrichment Using UMLS. Procedia Comput. Sci. 2013;23:78–83.

30. Rajput A. If It's on Web It's Yours! In: Pérez JB, Rodríguez JMC, Fähndrich J, Mathieu P, Campbell A, Suarez-Figueroa MC, et al., editors. Trends Pract. Appl. Agents Multiagent Syst. [Internet]. Springer International Publishing; 2013. p. 189–92. Available from: http://dx.doi.org/10.1007/978-3-319-00563-8_23

31. Matsuoka Y, Ghosh S, Kikuchi N, Kitano H. Payao: A community platform for SBML pathway model curation. Bioinformatics. 2010;26:1381–3.

32. Pennisi M, Rajput A-M, Toldo L, Pappalardo F. Agent based modeling of Treg-Teff cross regulation in relapsing-remitting multiple sclerosis. BMC Bioinformatics. 2013;14:S9.

33. Milo R, Kahana E, Milo R KE. Multiple sclerosis: geoepidemiology, genetics and the environment. Autoimmun Rev. 2010;9:A387-94.

34. Dua T, Rompani P, World Health Organization, Multiple Sclerosis International Federation. Atlas multiple sclerosis resources in the world, 2008. [Internet]. Geneva, Switzerland: World Health Organization; 2008 [cited 2014 Oct 30]. Available from: http://public.eblib.com/choice/publicfullrecord.aspx?p=380887

35. Goodin DS. Multiple Sclerosis and Related Disorders [Internet]. Amsterdam: Elsevier Science; 2014 [cited 2014 Dec 1]. Available from: http://public.eblib.com/choice/publicfullrecord.aspx?p=1629240

36. Clanet M. Jean-Martin Charcot. 1825 to 1893. Int. MS J. MS Forum. 2008;15:59–61.

37. Focus Shifts to Gray Matter in Search for the Cause of Multiple Sclerosis [Internet]. [cited 2015 Jan 10]. Available from: http://www.scientificamerican.com/article/focus-shifts-to-gray-matter-search-for-cause-multiple-sclerosis/

38. Donati D, Jacobson S. Viruses and Multiple Sclerosis. 2002 [cited 2015 Jan 10]; Available from: http://www.ncbi.nlm.nih.gov/books/NBK2494/

39. McKelvey C. Epigenetics in MS: A Primer. Mult. Scler. Discov. Forum [Internet]. 2014 [cited 2015 Jan 10]; Available from: http://www.msdiscovery.org/news/news_synthesis/12656-epigenetics-ms-primer

40. Mao P, Reddy PH. Is multiple sclerosis a mitochondrial disease? Biochim. Biophys. Acta. 2010;1802:66–79.

41. Ascherio A, Munger KL, White R, Köchert K, Simon KC, Polman CH, et al. Vitamin D as an Early Predictor of Multiple Sclerosis Activity and Progression. JAMA Neurol. 2014;71:306.

42. Oksenberg JR, Baranzini SE, Sawcer S, Hauser SL. The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. Nat. Rev. Genet. 2008;9:516–26.

43. How is Multiple Sclerosis diagnosed? [Internet]. [cited 2015 Jan 10]. Available from: http://www.msfocus.org/diagnose-multiple-sclerosis.aspx

44. Compston A CA, Compston A, Coles A. Multiple sclerosis. Lancet. 2008;372:1502–17.

45. Alonso A HM, Alonso A, Hernán MA. Temporal trends in the incidence of multiple sclerosis: a systematic review. Neurology. 2008;71:129–35.

46. Compston A CA, Compston A, Coles A. Multiple sclerosis. Lancet. 2002;359:1221–31.

47. Vukusic S, Confavreux C. Pregnancy and multiple sclerosis: the children of PRIMS. Clin. Neurol. Neurosurg. 2006;108:266–70.

48. Sibley WA, Bamford CR, Clark K. Clinical viral infections and multiple sclerosis. Lancet. 1985;1:1313–5.

49. Andersen O, Lygner PE, Bergström T, Andersson M, Vahlne A. Viral infections trigger multiple sclerosis relapses: a prospective seroepidemiological study. J. Neurol. 1993;240:417–22.

50. Panitch HS. Influence of infection on exacerbations of multiple sclerosis. Ann. Neurol. 1994;36 Suppl:S25-28.

51. Edwards S, Zvartau M, Clarke H, Irving W, Blumhardt L. Clinical relapses and disease activity on magnetic resonance imaging associated with viral upper respiratory tract infections in multiple sclerosis. J. Neurol. Neurosurg. Psychiatry. 1998;64:736–41.

52. SE B, Baranzini SE. Revealing the genetic basis of multiple sclerosis: are we there yet? Curr. Opin. Genet. Dev. 2011;21:317–24.

53. Ramagopalan SV, Maugeri NJ, Handunnetthi L, Lincoln MR, Orton S-M, Dyment DA, et al. Expression of the Multiple Sclerosis-Associated MHC Class II Allele HLA-DRB1*1501 Is Regulated by Vitamin D. Roopenian DC, editor. PLoS Genet. 2009;5:e1000369.

54. Lucchinetti C, Brück W, Parisi J, Scheithauer B, Rodriguez M, Lassmann H. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. Ann. Neurol. 2000;47:707–17.

55. Barnett MH, Prineas JW. Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion. Ann. Neurol. 2004;55:458–68.

56. Ascherio A MK, Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part&nbsp;I: the role of infection. Ann. Neurol. 2007;61:288–99.

57. Genain CP, Cannella B, Hauser SL, Raine CS. Identification of autoantibodies associated with myelin damage in multiple sclerosis. Nat. Med. 1999;5:170–5.

58. Prineas JW. The neuropathology of multiple sclerosis. Handb. Clin. Neurol. 1985;3:213–57.

59. Kornek B, Storch MK, Weissert R, Wallstroem E, Stefferl A, Olsson T, et al. Multiple sclerosis and chronic autoimmune encephalomyelitis: a comparative quantitative study of axonal injury in active, inactive, and remyelinated lesions. Am. J. Pathol. 2000;157:267–76.

60. Hohlfeld R, Wekerle H. Autoimmune concepts of multiple sclerosis as a basis for selective immunotherapy: from pipe dreams to (therapeutic) pipelines. Proc. Natl. Acad. Sci. U. S. A. 2004;101 Suppl 2:14599–606.

61. Ferguson B, Matyszak MK, Esiri MM, Perry VH. Axonal damage in acute multiple sclerosis lesions. Brain J. Neurol. 1997;120 (Pt 3):393–9.

62. Trapp BD, Peterson J, Ransohoff RM, Rudick R, Mörk S, Bö L. Axonal transection in the lesions of multiple sclerosis. N. Engl. J. Med. 1998;338:278–85.

63. Rudick RA, Fisher E, Lee J-C, Simon J, Jacobs L. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS. Neurology. 1999;53:1698–1698.

64. Simon JH, Jacobs LD, Campion MK, Rudick RA, Cookfair DL, Herndon RM, et al. A longitudinal study of brain atrophy in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). Neurology. 1999;53:139–48.

65. Rapalino O, Lazarov-Spiegler O, Agranov E, Velan GJ, Yoles E, Fraidakis M, et al. Implantation of stimulated homologous macrophages results in partial recovery of paraplegic rats. Nat. Med. 1998;4:814–21.

66. Moalem G, Leibowitz-Amit R, Yoles E, Mor F, Cohen IR, Schwartz M. Autoimmune T cells protect neurons from secondary degeneration after central nervous system axotomy. Nat. Med. 1999;5:49–55.

67. Hendrix S, Nitsch R. The role of T helper cells in neuroprotection and regeneration. J. Neuroimmunol. 2007;184:100–12.

68. Lu D, Goussev A, Chen J, Pannu P, Li Y, Mahmood A, et al. Atorvastatin reduces neurological deficit and increases synaptogenesis, angiogenesis, and neuronal survival in rats subjected to traumatic brain injury. J. Neurotrauma. 2004;21:21–32.

69. Sicotte M, Tsatas O, Jeong SY, Cai C-Q, He Z, David S. Immunization with myelin or recombinant Nogo-66/MAG in alum promotes axon regeneration and sprouting after corticospinal tract lesions in the spinal cord. Mol. Cell. Neurosci. 2003;23:251–63.

70. Mantovani A, Sica A, Locati M. Macrophage polarization comes of age. Immunity. 2005;23:344–6.

71. Weber MS, Prod'homme T, Youssef S, Dunn SE, Rundle CD, Lee L, et al. Type II monocytes modulate T cell-mediated central nervous system autoimmune disease. Nat. Med. 2007;13:935–43.

72. Storch MK, Stefferl A, Brehm U, Weissert R, Wallström E, Kerschensteiner M, et al. Autoimmunity to myelin oligodendrocyte glycoprotein in rats mimics the spectrum of multiple sclerosis pathology. Brain Pathol. Zurich Switz. 1998;8:681–94.

73. O'connor KC, Bar-Or A, Hafler DA. The Neuroimmunology of Multiple Sclerosis: Possible Roles of T and B Lymphocytes in Immunopathogenesis. J. Clin. Immunol. 2001;21:81–92.

74. Pette M, Fujita K, Wilkinson D, Altmann DM, Trowsdale J, Giegerich G, et al. Myelin autoreactivity in multiple sclerosis: recognition of myelin basic protein in the context of HLA-DR2 products by T lymphocytes of multiple-sclerosis patients and healthy donors. Proc. Natl. Acad. Sci. 1990;87:7968–72.

75. O'Connor KC, Appel H, Bregoli L, Call ME, Catz I, Chan JA, et al. Antibodies from Inflamed Central Nervous System Tissue Recognize Myelin Oligodendrocyte Glycoprotein. J. Immunol. 2005;175:1974–82.

76. Krogsgaard M, Wucherpfennig KW, Canella B, Hansen BE, Svejgaard A, Pyrdol J, et al. Visualization of Myelin Basic Protein (Mbp) T Cell Epitopes in Multiple Sclerosis Lesions Using a Monoclonal Antibody Specific for the Human Histocompatibility Leukocyte Antigen (Hla)-Dr2–Mbp 85–99 Complex. J. Exp. Med. 2000;191:1395–412.

77. Kieseier BC, Wiendl H, Hemmer B, Hartung H-P. Treatment and treatment trials in multiple sclerosis. Curr. Opin. Neurol. 2007;20:286–93.

78. Constantinescu CS, Gran B. Multiple sclerosis: autoimmune associations in multiple sclerosis. Nat. Rev. Neurol. 2010;6:591–2.

79. Steinman L. Multiple sclerosis: a coordinated immunological attack against myelin in the central nervous system. Cell. 1996;85:299–302.

80. Noseworthy JH. Progress in determining the causes and treatment of multiple sclerosis. Nature. 1999;399:A40-47.

81. Lucchinetti C, Brück W, Noseworthy J. Multiple sclerosis: recent developments in neuropathology, pathogenesis, magnetic resonance imaging studies and treatment. Curr. Opin. Neurol. 2001;14:259–69.

82. Steinman L. Multiple sclerosis: a two-stage disease. Nat. Immunol. 2001;2:762–4.

83. Wucherpfennig KW, Strominger JL. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. Cell. 1995;80:695–705.

84. Viglietta V, Baecher-Allan C, Weiner HL, Hafler DA. Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis. JExpMed. 2004;199:971–9.

85. Yamamura T, Gran B, editors. Multiple Sclerosis Immunology [Internet]. New York, NY: Springer New York; 2013 [cited 2015 Dec 11]. Available from: http://link.springer.com/10.1007/978-1-4614-7953-6

86. Optic Neuritis Study Group. Multiple sclerosis risk after optic neuritis: final optic neuritis treatment trial follow-up. Arch. Neurol. 2008;65:727–32.

87. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. Neurology. 1996;46:907–11.

88. Relapsing Remitting (RRMS) | Multiple Sclerosis Society UK [Internet]. [cited 2014 Dec 1]. Available from: http://www.mssociety.org.uk/what-is-ms/types-of-ms/relapsing-remitting-rrms

89. Trapp BD, Ransohoff RM, Fisher E, Rudick RA. Neurodegeneration in Multiple Sclerosis: Relationship to Neurological Disability. The Neuroscientist. 1999;5:48–57.

90. Weinshenker BG. Natural history of multiple sclerosis. Ann. Neurol. 1994;36 Suppl:S6-11.

91. Noseworthy JH. Progress in determining the causes and treatment of multiple sclerosis. Nature. 1999;399:A40–7.

92. Hauser SL, Oksenberg JR. The Neurobiology of Multiple Sclerosis: Genes, Inflammation, and Neurodegeneration. Neuron. 2006;52:61–76.

93. Trapp BD, Nave K-A. Multiple Sclerosis: An Immune or Neurodegenerative Disorder? Annu. Rev. Neurosci. 2008;31:247–69.

94. Primary-progressive MS (PPMS) [Internet]. Natl. Mult. Scler. Soc. [cited 2015 Jan 4]. Available from: http://www.nationalmssociety.org/What-is-MS/Types-of-MS/Primary-progressive-MS

95. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin. Pharmacol. Ther. 2001;69:89–95.

96. Fainardi E, Granieri E, Tola MR, Melchiorri L, Vaghi L, Rizzo R, et al. Clinical and MRI disease activity in multiple sclerosis are associated with reciprocal fluctuations in serum and cerebrospinal fluid levels of soluble HLA class I molecules. J. Neuroimmunol. 2002;133:151–9.

97. Fainardi E, Rizzo R, Melchiorri L, Vaghi L, Castellazzi M, Marzola A, et al. Presence of detectable levels of soluble HLA-G molecules in CSF of relapsing-remitting multiple sclerosis: relationship with CSF soluble HLA-I and IL-10 concentrations and MRI findings. J. Neuroimmunol. 2003;142:149–58.

98. Fainardi E, Rizzo R, Melchiorri L, Castellazzi M, Paolino E, Tola MR, et al. Intrathecal synthesis of soluble HLA-G and HLA-I molecules are reciprocally associated to clinical and MRI activity in patients with multiple sclerosis. Mult. Scler. Houndmills Basingstoke Engl. 2006;12:2–12.

99. Polman CH, Reingold SC, Edan G, Filippi M, Hartung H-P, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria." Ann. Neurol. 2005;58:840–6.

100. Villar LM, García-Barragán N, Sádaba MC, Espiño M, Gómez-Rial J, Martínez-San Millán J, et al. Accuracy of CSF and MRI criteria for dissemination in space in the diagnosis of multiple sclerosis. J. Neurol. Sci. 2008;266:34–7.

101. Tintoré M, Rovira A, Río J, Tur C, Pelayo R, Nos C, et al. Do oligoclonal bands add information to MRI in first attacks of multiple sclerosis? Neurology. 2008;70:1079–83.

102. Villar LM, Sádaba MC, Roldán E, Masjuan J, González-Porqué P, Villarrubia N, et al. Intrathecal synthesis of oligoclonal IgM against myelin lipids predicts an aggressive disease course in MS. J. Clin. Invest. 2005;115:187–94.

103. Thangarajh M, Gomez-Rial J, Hedström AK, Hillert J, Alvarez-Cermeño JC, Masterman T, et al. Lipid-specific immunoglobulin M in CSF predicts adverse long-term outcome in multiple sclerosis. Mult. Scler. Houndmills Basingstoke Engl. 2008;14:1208–13.

104. Miller DH. Biomarkers and surrogate outcomes in neurodegenerative disease: lessons from multiple sclerosis. NeuroRx J. Am. Soc. Exp. Neurother. 2004;1:284–94.

105. Weiner HL, Stankiewicz JM. Multiple sclerosis diagnosis and therapy [Internet]. Chichester, West Sussex; Hoboken, NJ: Wiley-Blackwell; 2012 [cited 2014 Dec 1]. Available from: http://public.eblib.com/choice/publicfullrecord.aspx?p=860887

106. Villoslada P. Biomarkers for multiple sclerosis. Drug News Perspect. 2010;23:585.

107. Treating Primary-progressive MS [Internet]. Natl. Mult. Scler. Soc. [cited 2015 Jan 7]. Available from: http://www.nationalmssociety.org/What-is-MS/Types-of-MS/Primary-progressive-MS/Treating-Primary-Progressive-MS

108. Barten LJ, Allington DR, Procacci KA, Rivey MP. New approaches in the management of multiple sclerosis. Drug Des. Devel. Ther. 2010;4:343–66.

109. Gupta GP, Nguyen DX, Chiang AC, Bos PD, Kim JY, Nadal C, et al. Mediators of vascular remodelling co-opted for sequential steps in lung metastasis. Nature. 2007;446:765–70.

110. What is systems biology? : The Seven Stones [Internet]. [cited 2015 Jan 11]. Available from: http://blogs.nature.com/sevenstones/2007/07/what_is_systems_biology_3.html

111. Kitano H. Computational systems biology. Nature. 2002;420:206–10.

112. Zak DE. Systems biology of innate immunity. Immunol Rev. 2009;227:264–82.

113. Boran ADW, Iyengar R. Systems approaches to polypharmacology and drug discovery. Curr. Opin. Drug Discov. Devel. 2010;13:297–309.

114. Kitano H. Foundations of systems biology. Cambridge, Mass.: MIT Press; 2001.

115. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860–921.

116. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein–protein interaction network. Nature. 2005;437:1173–8.

117. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, et al. A human protein-protein interaction network: a resource for annotating the proteome. Cell. 2005;122:957–68.

118. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, et al. Large-scale mapping of human protein–protein interactions by mass spectrometry. Mol. Syst. Biol. 2007;3:89.

119. Scriver CR, Waters PJ. Monogenic traits are not simple: lessons from phenylketonuria. Trends Genet. TIG. 1999;15:267–72.

120. Dipple KM, McCabe ER. Modifier genes convert “simple” Mendelian disorders to complex traits. Mol. Genet. Metab. 2000;71:43–50.

121. Pache RA, Zanzoni A, Naval J, Mas JM, Aloy P. Towards a molecular characterisation of pathological pathways. FEBS Lett. 2008;582:1259–65.

122. Cohen AA, Geva-Zatorsky N, Eden E, Frenkel-Morgenstern M, Issaeva I, Sigal A, et al. Dynamic Proteomics of Individual Cancer Cells in Response to a Drug. Science. 2008;322:1511–6.

123. Zhang Z, Larner SF, Kobeissy F, Hayes RL, Wang KKW. Systems Biology and Theranostic Approach to Drug Discovery and Development to Treat Traumatic Brain Injury. In: Yan Q, editor. Syst. Biol. Drug Discov. Dev. [Internet]. Totowa, NJ: Humana Press; 2010 [cited 2014 Nov 27]. p. 317–29. Available from: http://link.springer.com/10.1007/978-1-60761-800-3_16

124. Noble D, Garny A, Noble PJ. How the Hodgkin-Huxley equations inspired the Cardiac Physiome Project. J. Physiol. 2012;590:2613–28.

125. Bakker BM, Michels PA, Opperdoes FR, Westerhoff HV. Glycolysis in bloodstream form Trypanosoma brucei can be understood in terms of the kinetics of the glycolytic enzymes. J. Biol. Chem. 1997;272:3207–15.

126. Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, et al. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. Eur. J. Biochem. FEBS. 2000;267:5313–29.

127. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.

128. Karp PD, Riley M, Saier M, Paulsen IT, Paley SM, Pellegrini-Toole A. The EcoCyc and MetaCyc databases. Nucleic Acids Res. 2000;28:56–9.

129. Loomis WF, Sternberg PW. Genetic networks. Science. 1995;269:649.

130. Kell DB. Metabolomics and systems biology: making sense of the soup. Curr. Opin. Microbiol. 2004;7:296–307.

131. Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Annu. Rev. Genomics Hum. Genet. 2001;2:343–72.

132. Schneider H-C, Klabunde T. Understanding drugs and diseases by systems biology? Bioorg. Med. Chem. Lett. 2013;23:1168–76.

133. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. Nat. Genet. 1999;21:33–7.

134. DeRisi JL, Iyer VR, Brown PO. Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. Science. 1997;278:680–6.

135. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, et al. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell. 1998;9:3273–97.

136. Casadesus G, Arendash G, Laferla F, McDonald M, Casadesus G, Arendash G, et al. Animal Models of Alzheimer's Disease, Animal Models of Alzheimer's Disease. Int. J. Alzheimer's Dis. Int. J. Alzheimer's Dis. 2011;2010, 2010:e606357.

137. A Comprehensive Molecular Interaction Map for Rheumatoid Arthritis. PLoS ONE. 2010;5:e10137.

138. Li J, Zhu X, Chen JY. Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts. PLoS Comput. Biol. 2009;5:e1000450.

139. Jitendra S, Nanda A, Kaur S, Singh M. A comprehensive molecular interaction map for Hepatitis B virus and drug designing of a novel inhibitor for Hepatitis B X protein. Bioinformation. 2011;7:9–14.

140. Small DH, San Mok S, Bornstein JC. Alzheimer's disease and Aβ toxicity: from top to bottom. Nat. Rev. Neurosci. 2001;2:595–598.

141. Jensen LJ, Saric J, Bork P. Literature mining for the biologist: from information retrieval to biological discovery. Nat. Rev. Genet. 2006;7:119–29.

142. Information NC for B, Pike USNL of M 8600 R, MD B, Usa 20894. PubMed Help. 2014 [cited 2015 Feb 2]; Available from: http://www.ncbi.nlm.nih.gov/books/NBK3827/

143. Agrawal DS, Rao PMRS. Biological Disease Mechanism Networks. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, editors. Encycl. Syst. Biol. [Internet]. Springer New York; 2013 [cited 2015 Jan 14]. p. 113–8. Available from: http://link.springer.com/referenceworkentry/10.1007/978-1-4419-9863-7_583

144. Rosen R. Systems Theory and Biology. Proceedings of the 3rd Systems Symposium, Cleveland, Ohio, Oct. 1966. M. D. Mesarović, Ed. Springer-Verlag, New York, 1968. xii 403 pp., illus. $16. Science. 1968;161:34–5.

145. THOM R. Stabilite Structurelle et Morphogenese [Internet]. [cited 2015 Jan 21]. Available from: http://www.biblio.com/book/stabilite-structurelle-morphogenese-thom-r/d/638483592

146. Rosen R. Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life. Auflage: New Ed. New York: Columbia Univ Pr; 2005.

147. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat. Biotechnol. 2010;28:245–8.

148. Arkin A, Ross J, McAdams HH. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. Genetics. 1998;149:1633–48.

149. Gilbert J, Henske P, Singh A. Rebuilding big pharma's business model. VIVO-N. Y. THEN NORWALK-. 2003;21:73–80.

150. Materi W, Wishart DS. Computational systems biology in drug discovery and development: methods and applications. Drug Discov. Today. 2007;12:295–303.

151. Aradi I, Erdi P. Computational neuropharmacology: dynamical approaches in drug discovery. Trends Pharmacol. Sci. 2006;27:240–3.

152. Baranzini SE, Bernard CCA, Oksenberg JR. Modular transcriptional activity characterizes the initiation and progression of autoimmune encephalomyelitis. J. Immunol. Baltim. Md 1950. 2005;174:7412–22.

153. Woolfe F, Waxman SG, Hains BC. In silico modeling of axonal reconnection within a discrete fiber tract after spinal cord injury. J. Neurotrauma. 2007;24:421–32.

154. T1DMS [Internet]. Epsil. Group. [cited 2015 Jan 29]. Available from: https://tegvirginia.com/solutions/t1dms/

155. Qualitative reasoning [Internet]. Wikipedia Free Encycl. 2015 [cited 2015 Jan 27]. Available from: http://en.wikipedia.org/w/index.php?title=Qualitative_reasoning&oldid=619374011

156. Heiner M, Donaldson R, Gilbert D. Petri nets for systems biology. Symb. Syst. Biol. Theory Methods Jones Bartlett Publ. Inc USA Press 2010 [Internet]. 2010 [cited 2015 Jan 29]; Available from: http://www.stanford.edu/~radonald/index_files/jones%202008.pdf

157. Interactome [Internet]. Wikipedia Free Encycl. 2016 [cited 2016 Apr 21]. Available from: https://en.wikipedia.org/w/index.php?title=Interactome&oldid=713458417

158. Protein–protein interaction [Internet]. Wikipedia Free Encycl. 2016 [cited 2016 Apr 21]. Available from: https://en.wikipedia.org/w/index.php?title=Protein%E2%80%93protein_interaction&oldid=713402377

159. Gene regulatory network [Internet]. Wikipedia Free Encycl. 2016 [cited 2016 Apr 21]. Available from: https://en.wikipedia.org/w/index.php?title=Gene_regulatory_network&oldid=715195996

160. Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Annu. Rev. Genomics Hum. Genet. 2001;2:343–372.

161. BioGRID Database Statistics | BioGRID [Internet]. [cited 2015 Jan 14]. Available from: http://wiki.thebiogrid.org/doku.php/statistics

162. Mizuno S, Iijima R, Ogishima S, Kikuchi M, Matsuoka Y, Ghosh S, et al. AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. BMC Syst. Biol. 2012;6:52.

163. Kitano H, Oda K, Kimura T, Matsuoka Y, Csete M, Doyle J, et al. Metabolic Syndrome and Robustness Tradeoffs. Diabetes. 2004;53:S6–15.

164. Oda K, Kimura T, Matsuoka Y, Funahashi A, Muramatsu M, Kitano H. Molecular interaction map of a macrophage. AfCS Res. Rep. 2004;2:1–12.

165. Fujita KA, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, et al. Integrating pathways of Parkinson's disease in a molecular interaction map. Mol. Neurobiol. 2014;49:88–102.

166. Tortolina L, Castagnino N, De Ambrosi C, Moran E, Patrone F, Ballestrero A, et al. A multi-scale approach to colorectal cancer: from a biochemical- interaction signaling-network level, to multi-cellular dynamics of malignant transformation. Interplay with mutations and onco-protein inhibitor drugs. Curr. Cancer Drug Targets. 2012;12:339–55.

167. Le Novere N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The systems biology graphical notation. Nat. Biotechnol. 2009;27:735–741.

168. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 2003;19:524–31.

169. Main Page - SBML.caltech.edu [Internet]. [cited 2015 Oct 27]. Available from: http://sbml.org/Main_Page

170. BioModels Database [Internet]. [cited 2015 Oct 27]. Available from: http://www.ebi.ac.uk/biomodels-main/

171. Le Novere N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The Systems Biology Graphical Notation. Nat Biotechnol. 2009;27:735–41.

172. Wu G, Zhu L, Dent JE, Nardini C. A Comprehensive Molecular Interaction Map for Rheumatoid Arthritis. PLoS ONE. 2010;5:e10137.

173. Systems Biology Graphical Notation [Internet]. Wikipedia. 2017 [cited 2017 Nov 5]. Available from: https://en.wikipedia.org/w/index.php?title=Systems_Biology_Graphical_Notation&oldid=800739626

174. Reactome Pathway Database [Internet]. [cited 2015 Nov 10]. Available from: http://www.reactome.org/

175. BioCyc Pathway/Genome Database Collection [Internet]. [cited 2015 Nov 10]. Available from: http://biocyc.org/

176. Pathway Commons: A Resource for Biological Pathway Analysis [Internet]. [cited 2015 Nov 10]. Available from: http://www.pathwaycommons.org/about/

177. WikiPathways - WikiPathways [Internet]. [cited 2015 Nov 10]. Available from: http://www.wikipathways.org/index.php/WikiPathways

178. Demir E, Babur Ö, Rodchenkov I, Aksoy BA, Fukuda KI, Gross B, et al. Using Biological Pathway Data with Paxtools. PLoS Comput Biol. 2013;9:e1003194.

179. Systems Biology Linker (Sybil) | Reviews for Systems Biology Linker (Sybil) at SourceForge.net [Internet]. [cited 2015 Nov 10]. Available from: http://sourceforge.net/projects/sbpax/reviews

180. Babur O, Dogrusoz U, Demir E, Sander C. ChiBE: interactive visualization and manipulation of BioPAX pathway models. Bioinforma. Oxf. Engl. 2010;26:429–31.

181. Rodchenkov I, Demir E, Sander C, Bader GD. The BioPAX Validator. Bioinforma. Oxf. Engl. 2013;29:2659–60.

182. Schlage WK, Westra JW, Gebel S, Catlett NL, Mathis C, Frushour BP, et al. A computable cellular stress network model for non-diseased pulmonary and cardiovascular tissue. BMC Syst. Biol. 2011;5:168.

183. Walker DC, Southgate J. The virtual cell—a candidate co-ordinator for ‘middle-out’ modelling of biological systems. Brief. Bioinform. 2009;10:450–61.

184. Wang Z, Deisboeck TS. Computational modeling of brain tumors: discrete, continuum or hybrid? Sci. Model. Simul. SMNS. 2008;15:381–93.

185. Greene JM, Levy D, Fung KL, Silva de Souza P, Gottesman MM, Lavi O. Modeling intrinsic heterogeneity and growth of cancer cells. J. Theor. Biol. 2014;

186. De Schutter E. Why are computational neuroscience and systems biology so separate? PLoS Comput. Biol. 2008;4:e1000078.

187. Mizuno S, Iijima R, Ogishima S, Kikuchi M, Matsuoka Y, Ghosh S, et al. AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. BMC Syst. Biol. 2012;6:52.

188. Fujita KA, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, et al. Integrating Pathways of Parkinson's Disease in a Molecular Interaction Map. Mol. Neurobiol. 2014;49:88–102.

189. Simmons SB, Pierson ER, Lee SY, Goverman JM. Modeling the heterogeneity of multiple sclerosis in animals. Trends Immunol. 2013;34:410–22.

190. Μικελλίδης Λ, Mikellides L. Modelling the multiple sclerosis disease using stochastic petri nets. https://ktree.cs.ucy.ac.cy/action.php?kt_path_info=ktcore.actions.document.view&fDocumentId=12886 [Internet]. 2011 [cited 2016 Nov 15]; Available from: http://lekythos.library.ucy.ac.cy/handle/10797/13112

191. Hellriegel B, Daumer M, Neiss A. Analysing the course of multiple sclerosis with segmented regression models [Internet]. Discussion paper//Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München; 2003. Available from: http://www.econstor.eu/handle/10419/31091

192. Heitjan DF. Nonlinear Modeling of Serial Immunologic Data: A Case Study. J. Am. Stat. Assoc. 1991;86:891–8.

193. Albert PS, McFarland HF, Smith ME, Frank JA. Time series for modelling counts from a relapsing-remitting disease: application to modelling disease activity in multiple sclerosis. Stat. Med. 1994;13:453–66.

194. Motl RW, Mullen S, Suh Y, McAuley E. Does physical activity change over 24 months in persons with relapsing-remitting multiple sclerosis? Health Psychol. Off. J. Div. Health Psychol. Am. Psychol. Assoc. 2014;33:326–31.

195. Le Teuff G, Abrahamowicz M, Wynant W, Binquet C, Moreau T, Quantin C. Flexible modeling of disease activity measures improved prognosis of disability progression in relapsing-remitting multiple sclerosis. J. Clin. Epidemiol. 2015;68:307–16.

196. Gauthier SA, Mandel M, Guttmann CRG, Glanz BI, Khoury SJ, Betensky RA, et al. Predicting short-term disability in multiple sclerosis. Neurology. 2007;68:2059–65.

197. Palace J, Bregenzer T, Tremlett H, Oger J, Zhu F, Zhu F, et al. UK multiple sclerosis risk-sharing scheme: a new natural history dataset and an improved Markov model. BMJ Open. 2014;4:e004073.

198. Di Serio C, Lamina C, Di Serio C, Lamina C. Investigating Determinants of Multiple Sclerosis in Longitunal Studies: A Bayesian Approach, Investigating Determinants of Multiple

Sclerosis in Longitunal Studies: A Bayesian Approach. J. Probab. Stat. J. Probab. Stat. 2009;2009, 2009:e198320.

199. Lawton M, Tilling K, Robertson N, Tremlett H, Zhu F, Harding K, et al. A longitudinal model for disease progression was developed and applied to multiple sclerosis. J. Clin. Epidemiol. 2015;68:1355–65.

200. Nixon R, Bergvall N, Tomic D, Sfikas N, Cutter G, Giovannoni G. No Evidence of Disease Activity: Indirect Comparisons of Oral Therapies for the Treatment of Relapsing–Remitting Multiple Sclerosis. Adv. Ther. 2014;31:1134–54.

201. Pennisi M, Rajput A-M, Toldo L, Pappalardo F. Agent based modeling of Treg-Teff cross regulation in relapsing-remitting multiple sclerosis. BMC Bioinformatics. 2013;14:S9.

202. Pappalardo F, Pennisi M, Rajput A-M, Chiacchio F, Motta S. Relapsing-remitting Multiple Scleroris and the Role of Vitamin D: An Agent Based Model. Proc. 5th ACM Conf. Bioinforma. Comput. Biol. Health Inform. [Internet]. New York, NY, USA: ACM; 2014 [cited 2015 Dec 9]. p. 744–748. Available from: http://doi.acm.org/10.1145/2649387.2660844

203. Pennisi M, Russo G, Motta S, Pappalardo F. Agent based modeling of the effects of potential treatments over the blood-brain barrier in multiple sclerosis. J. Immunol. Methods. 2015;

204. Kuceyeski AF, Vargas W, Dayan M, Monohan E, Blackwell C, Raj A, et al. Modeling the relationship among gray matter atrophy, abnormalities in connecting white matter, and cognitive performance in early multiple sclerosis. AJNR Am. J. Neuroradiol. 2015;36:702–9.

205. Meier DS, Guttmann CRG. MRI time series modeling of MS lesion development. NeuroImage. 2006;32:531–7.

206. Ensari I, Motl RW, McAuley E, Mullen SP, Feinstein A. Patterns and predictors of naturally occurring change in depressive symptoms over a 30-month period in multiple sclerosis. Mult. Scler. Houndmills Basingstoke Engl. 2014;20:602–9.

207. Basic Formal Ontology (BFO) | Home [Internet]. [cited 2016 May 31]. Available from: http://ifomis.uni-saarland.de/bfo/

208. OntoFast: Construct Ontology Rapidly - Springer [Internet]. [cited 2014 Dec 11]. Available from: http://link.springer.com/chapter/10.1007%2F978-3-319-11716-4_21

209. Fernández-Suárez XM, Rigden DJ, Galperin MY. The 2014 Nucleic Acids Research Database Issue and an updated NAR online Molecular Biology Database Collection. Nucleic Acids Res. 2014;42:D1–6.

210. Malhotra A, Younesi E, Gündel M, Müller B, Heneka MT, Hofmann-Apitius M. ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease. Alzheimers Dement. 2014;10:238–46.

211. Fraunhofer SCAI: Downloads [Internet]. [cited 2015 Feb 10]. Available from: http://www.scai.fraunhofer.de/en/business-research-areas/bioinformatics/downloads.html

212. Multiple Sclerosis Encyclopedia/Glossary [Internet]. [cited 2014 Aug 21]. Available from: http://www.mult-sclerosis.org/wholeglossary.html

213. Vocabularies in the UMLS® Metathesaurus [Internet]. [cited 2015 Feb 3]. Available from: http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/source_faq.html

214. Load Scripts [Internet]. 2013 [cited 2013 May 6]. Available from: http://www.nlm.nih.gov/research/umls/implementation_resources/scripts/index.html

215. UMLS - Semantic Network MYSQL Load Script [Internet]. [cited 2017 Nov 4]. Available from: https://www.nlm.nih.gov/research/umls/implementation_resources/scripts/README_Semantic_Network_MySQL.html

216. Sample UMLS Metathesaurus in Microsoft Access® [Internet]. 2013 [cited 2013 May 7]. Available from: http://www.nlm.nih.gov/research/umls/implementation_resources/community/dbloadscripts/ms_access.html

217. Taverna - open source and domain independent Workflow Management System [Internet]. 2013 [cited 2013 May 7]. Available from: http://www.taverna.org.uk/

218. MySQL :: Download Connector/J [Internet]. [cited 2017 Oct 21]. Available from: https://dev.mysql.com/downloads/connector/j/5.1.html

219. Welcome to the NCBO BioPortal | NCBO BioPortal [Internet]. 2014 [cited 2014 May 16]. Available from: http://bioportal.bioontology.org/

220. Espinoza M, Gómez-Pérez A, Mena E. Labeltranslator-a tool to automatically localize an ontology. Semantic Web Res. Appl. [Internet]. Springer; 2008 [cited 2014 May 15]. p. 792–796. Available from: http://link.springer.com/chapter/10.1007/978-3-540-68234-9_60

221. Wählen Sie den unabhängigen Browser [Internet]. Mozilla. [cited 2016 Jun 1]. Available from: https://www.mozilla.org/de/firefox/new/

222. Chrome-Browser [Internet]. [cited 2016 Jun 1]. Available from: https://www.google.de/chrome/browser/desktop/

223. DownThemAll! [Internet]. [cited 2016 Sep 12]. Available from: http://www.downthemall.net/

224. Link Gopher [Internet]. [cited 2016 Jun 1]. Available from: https://addons.mozilla.org/de/firefox/addon/link-gopher/

225. gnu.org [Internet]. [cited 2016 Jun 1]. Available from: https://www.gnu.org/software/grep/

226. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

227. Novère NL, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The Systems Biology Graphical Notation. Nat. Biotechnol. 2009;27:735–41.

228. DBGET Search [Internet]. [cited 2017 Oct 20]. Available from: http://www.genome.jp/dbget/

229. Saccharomyces Genome Database | SGD [Internet]. [cited 2017 Oct 20]. Available from: https://www.yeastgenome.org/

230. iHOP - Information Hyperlinked over Proteins [Internet]. [cited 2017 Oct 20]. Available from: http://www.ihop-net.org/UniPub/iHOP/

231. Genome Network Platform [Internet]. [cited 2017 Oct 20]. Available from: http://genomenetwork.nig.ac.jp/

232. pubmeddev. Home - PubMed - NCBI [Internet]. [cited 2017 Oct 20]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/

233. Home - Gene - NCBI [Internet]. [cited 2017 Oct 20]. Available from: https://www.ncbi.nlm.nih.gov/gene

234. SABIO Biochemical Reaction Kinetics Database [Internet]. [cited 2017 Oct 20]. Available from: http://sabio.h-its.org/

235. Chelliah V, Juty N, Ajmera I, Ali R, Dumousseau M, Glont M, et al. BioModels: ten-year anniversary. Nucleic Acids Res. 2015;43:D542–8.

236. Kutmon M, van Iersel MP, Bohler A, Kelder T, Nunes N, Pico AR, et al. PathVisio 3: An Extendable Pathway Analysis Toolbox. Murphy RF, editor. PLOS Comput. Biol. 2015;11:e1004085.

237. Sauro HM, Hucka M, Finney A, Wellock C, Bolouri H, Doyle J, et al. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. Omics J. Integr. Biol. 2003;7:355–72.

238. MedDRA [Internet]. [cited 2015 Mar 4]. Available from: http://www.meddra.org/

239. Payao beta [Internet]. [cited 2014 Sep 24]. Available from: http://sblab.celldesigner.org/Payao10/bin/

240. Vosoughi R, Freedman MS. Therapy of MS. Clin. Neurol. Neurosurg. 2010;112:365–85.

241. Fontoura P, Garren H. Multiple sclerosis therapies: molecular mechanisms and future. Results Probl. Cell Differ. 2010;51:259–85.

242. Ramagopalan SV, Dobson R, Meier UC, Giovannoni G. Multiple sclerosis: risk factors, prodromes, and potential causal pathways. Lancet Neurol. 2010;9:727–39.

243. Pharmaceutical clinical trial information tracking [Internet]. [cited 2015 Feb 9]. Available from: http://www.citeline.com/products/trialtrove/

244. Home : National Multiple Sclerosis Society [Internet]. [cited 2014 Dec 11]. Available from: http://www.nationalmssociety.org/

245. DrugBank - the Datahub [Internet]. [cited 2016 Jun 1]. Available from: https://datahub.io/dataset/fu-berlin-drugbank

246. TEMIS [Internet]. [cited 2016 Jun 1]. Available from: http://www.temis.com/index.en.html

247. Zotero | Home [Internet]. [cited 2015 May 16]. Available from: https://www.zotero.org/

248. Guo Z, Han HK, Tay JC. Sufficiency verification of HIV-1 pathogenesis based on multi-agent simulation. Proc ACM Genet. Evol. Comput. Conf. GECCO05. 2005;305–12.

249. Perrin D, Ruskin HJ, Burns J, Crane M. An agent-based approach to immune modelling. Lect. Notes Comput. Sci. 2006;3980:612–21.

250. Bernaschi M, Castiglione F. Design and implementation of an immune system simulator. Comp Biol Med. 2001;3:303–31.

251. Pappalardo F, Lollini P-L, Castiglione F, Motta S. Modelling and simulation of cancer immuno-prevention vaccine. Bioinformatics. 2005;21:2891–7.

252. Pennisi M, Pappalardo M, Palladini A, Nicoletti G, Nanni P, Lollini P-L, et al. Modeling the competition between lung metastases and the immune system using agents. BMC Bioinformatics. 2010;11:S13.

253. Pappalardo F, Cincotti A, Motta S, Pennisi M. Agent based modeling of atherosclerosis: a concrete help in personalized treatments. Lect. Notes Artif. Intell. 2009;5755:386–96.

254. Palladini A, Nicoletti G, Pappalardo F, Murgo A, Grosso V, Stivani V, et al. In silico modeling and in vivo efficacy of cancer preventive vaccinations. Cancer Res. 2010;70:7755–63.

255. Sakaguchi S, Sakaguchi N, Asano M, Itoh M, Toda M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alphachains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. J Immunol. 1995;155:1151–64.

256. Thornton AM, Shevach EM. Suppressor effector function of CD4+CD25+ immunoregulatory T cells is antigen nonspecific. J Immunol. 2000;164:183–90.

257. Shevach EM, McHugh RS, Piccirillo CA, Thornton AM. Control of T-cell activation by CD4+ CD25+ suppressor T cells. Immunol Rev. 2001;182:58–67.

258. Fontenot JD, Rudensky AY. A well adapted regulatory contrivance: regulatory T cell development and the forkhead family transcription factor Foxp3. Nat Immunol. 2005;6:331–7.

259. Van der Vliet HJ, Nieuwenhuis EE. IPEX as a result of mutations in FOXP3. Clin Dev Immunol. 2007;2007:89017.

260. Kukreja A, Cost G, Marker J, Zhang C, Sun Z, Lin-Su K, et al. Multiple immuno-regulatory defects in type-1 diabetes. J Clin Invest. 2002;109:131–40.

261. Elizabeth S. NetLogo, a multi-agent simulation environment. Artif. Life. 2011;13:303–11.

262. Sundstrom P, Juto P, Wadell G, Hallmans G, Svenningsson A, Nystrom L, et al. An altered immune response to Epstein-Barr virus in multiple sclerosis A prospective study. NEUROLOGY. 2004;62:2277–82.

263. Sospedra M, Martin R. Immunology of multiple sclerosis. Annu Rev Immunol. 2005;23:683–747.

264. Ivchenko O, Younesi E, Shahid M, Wolf A, Müller B, Hofmann-Apitius M. PLIO: an ontology for formal description of protein–ligand interactions. Bioinformatics. 2011;27:1684–90.

265. Apache Nutch^{TM} - [Internet]. [cited 2015 Mar 28]. Available from: http://nutch.apache.org/

266. WebSPHINX: A Personal, Customizable Web Crawler [Internet]. [cited 2015 Mar 28]. Available from: http://www.cs.cmu.edu/~rcm/websphinx/

267. OpenWebSpider [Internet]. [cited 2015 Mar 28]. Available from: http://www.openwebspider.org/

268. LinkedCT Databrowse [Internet]. [cited 2014 Dec 11]. Available from: http://linkedct.org/

269. Hoffmann R. A wiki for the life sciences where authorship matters. Nat. Genet. 2008;40:1047–51.

270. Mons B, Ashburner M, Chichester C, van Mulligen E, Weeber M, den Dunnen J, et al. Calling on a million minds for community annotation in WikiProteins. Genome Biol. 2008;9:R89.

271. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. PLoS Biol. 2008;6:e184.

272. Muljadi H, Takeda H, Shakya A, Kawamoto S, Kobayashi S, Fujiyama A, et al. Semantic Wiki as a Lightweight Knowledge Management System. Semantic Web - ASWC 2006 First Asian Semantic Web Conf. Beijing China Sept. 3-7 2006 Proc. [Internet]. 2006. p. 65–71. Available from: http://dx.doi.org/10.1007/11836025_7

273. Semantic Wikis: A Comprehensible Introduction with Examples from the Health Sciences | Boulos | Journal of Emerging Technologies in Web Intelligence [Internet]. [cited 2015 May 16]. Available from: http://ojs.academypublisher.com/index.php/jetwi/article/view/01019496

274. sbml2smw - CellDesigner Plugin that stores CellDesigner models in a Semantic MediaWiki - Google Project Hosting [Internet]. [cited 2014 Dec 11]. Available from: http://code.google.com/p/sbml2smw/

275. Van Hemert JL, Dickerson JA. PathwayAccess: CellDesigner plugins for pathway databases. Bioinforma. Oxf. Engl. 2010;26:2345–6.

276. Help:Start [Internet]. [cited 2015 May 16]. Available from: http://iwb.fluidops.com/resource/Help:Start?view=wiki

277. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, et al. COPASI—a COmplex PAthway SImulator. Bioinformatics. 2006;22:3067–74.

278. larkc [Internet]. [cited 2014 Aug 21]. Available from: http://pipes.yahoo.com/pipes/pipe.run?_id=3e447f6fc4b9d447447ede5790d7b140&_render=rss

279. LinkedOpenData-LOD [Internet]. [cited 2014 Aug 21]. Available from: http://richard.cyganiak.de/2007/10/lod/imagemap.html

280. Weinstock-Guttman B, Bakshi R. Combination therapy for multiple sclerosis: the treatment strategy of the future? CNS Drugs. 2004;18:777–92.

281. Stuart WH. Combination therapy for the treatment of multiple sclerosis: challenges and opportunities. Curr. Med. Res. Opin. 2007;23:1199–208.

282. Conway D, Cohen JA. Combination therapy in multiple sclerosis. Lancet Neurol. 2010;9:299–308.

283. Lublin FD, Cofield SS, Cutter GR, Conwit R, Narayana PA, Nelson F, et al. Randomized study combining interferon and glatiramer acetate in multiple sclerosis. Ann. Neurol. 2013;73:327–40.

284. Trapp BD, Ransohoff R, Rudick R. Axonal pathology in multiple sclerosis: relationship to neurologic disability. Curr. Opin. Neurol. 1999;12:295–302.

285. Mendizábal NV de, Carneiro J, Solé RV, Goñi J, Bragard J, Martinez-Forero I, et al. Modeling the effector - regulatory T cell cross-regulation reveals the intrinsic character of relapses in Multiple Sclerosis. BMC Syst. Biol. 2011;5:1.

286. Boiko A, Vorobeychik G, Paty D, Devonshire V, Sadovnick D, University of British Columbia MS Clinic Neurologists. Early onset multiple sclerosis: a longitudinal study. Neurology. 2002;59:1006–10.

287. Steinman L. Multiple sclerosis: a two-stage disease. Nat. Immunol. 2001;2:762–4.