IDENTIFICATION OF PROTEIN-PROTEIN INTERACTION BRIDGES FOR MULTIPLE SCLEROSIS

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

COMPUTER ENGINEERING

By

Gözde Yazıcı

December 2022

Identification of protein-protein interaction bridges for multiple sclerosis
By Gözde Yazıcı
December 2022

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Can Alkan(Advisor)

Uğur Doğrusöz

Tunca Doğan

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

IDENTIFICATION OF PROTEIN-PROTEIN INTERACTION BRIDGES FOR MULTIPLE SCLEROSIS

Gözde Yazıcı
M.S. in Computer Engineering
Advisor: Can Alkan
December 2022

Identifying and prioritizing disease-related proteins is an important scientific problem to understand disease etiology. Network science has become an important discipline to prioritize such proteins. Multiple sclerosis (MS), an autoimmune disease which still cannot be cured, is characterized by a damaging process called demyelination. Demyelination is the destruction of the crucial nerve sheath, myelin, and oligodendrocytes, the cells producing myelin, by immune cells. Identifying the proteins having special features on the network formed by the proteins of oligodendrocyte and immune cells can reveal useful information about the disease. To this end, we investigated the most significant protein pairs for the intra- and intercellular protein networks that we define as bridges among the proteins providing the interaction between the two cells in demyelination. We analyzed two protein networks including the oligodendrocyte and each type of two immune cells, macrophage and T-cell. We developed a model called BriFin that prioritizes contact protein pairs using network analysis techniques and integer programming. We showed several proteins it prioritized have already been associated with MS in the relevant literature. For the oligodendrocyte-macrophage network, we showed that 77% to 100% of the proteins BriFin detected, depending on the parametrization, are MS-associated. We further experimentally investigated 4 proteins prioritized by BriFin, and observed that the mRNA expression levels of 2 out of these 4 proteins significantly decreased in a group of MS patients. We therefore here present a model, BriFin, which can be used to analyze processes where interactions of two cell types play an important role.

Keywords: Multiple Sclerosis, Demyelination, Integer Programming, Network Analysis, Protein-protein Interaction.

ÖZET

MULTİPL SKLEROZ İÇİN PROTEİN-PROTEİN ETKİLEŞİM KÖPRÜLERİNİN BELİRLENMESİ

Gözde Yazıcı
Bilgisayar Mühendisliği, Yüksek Lisans
Tez Danışmanı: Can Alkan
Aralık 2022

Bir hastalığın nedenlerini anlayabilmek için hastalıkla ilişkili proteinlerin belirlenmesi ve önceliklendirilmesi önemli bir bilimsel problemdir. Ağ (network) bilimi bu proteinlerin önceliklendirilmesi için önemli bir disiplin haline gelmiştir. Halen tamamen tedavi edilemeyen bir otoimmün hastalık olan multipl skleroz (MS) demiyelinizasyon olarak adlandırılan tahrip edici bir biyolojik olay ile karakterize edilir. Demiyelinizasyon, yüksek önem taşıyan sinir kılıfı miyelinin ve miyelini üreten hücreler olan oligodendrositlerin savunma hücreleri tarafından tahrip edilmesidir. Oligodendrosit ve savunma hücrelerinin proteinlerinden oluşan protein ağında özel nitelikleri bulunan proteinlerin tespit edilmesi, hastalık ile ilgili faydalı bilgiler açığa çıkarabilir. Bu amaçla, demiyelinizasyonda rol alan iki hücrenin etkileşimini sağlayan proteinler arasından köprüler olarak tanımladığımız hücre içi ve hücreler arası protein ağları için en önemli protein çiftlerini araştırdık. Oligodendrosit ile makrofaj ve T hücresi olmak üzere iki tip savunma hücresinin oluşturduğu iki protein ağını analiz ettik. Hücreler arası etkileşim sağlayan proteinleri ağ analiz teknikleri ve tam sayılı programlama ile önceliklendiren BriFin olarak adlandırdığımız bir metot geliştirdik. BriFin'in öncelikli olarak belirlediği proteinlerin bir kısmının ilişkili literatürde hastalık ile hâlihazırda ilişkilendirildiğini gösterdik. Oligodendrosit-makrofaj ağı için, BriFin'in tespit ettiği proteinlerin parametrizasyona bağlı olarak %77 ila %100'ünün MS ile ilişkili olduğunu gösterdik. Ayrıca, BriFin ile önceliklendirdiğimiz 4 proteini deneysel olarak araştırarak 2 tanesinin mRNA ekspresyon seviyelerinin bir grup MS hastasında anlamlı ölçüde azaldığını gözlemledik. Dolayısıyla burada, iki hücre arasındaki etkileşimlerin önemli rol oynadığı süreçleri analiz etmek için kullanılabilecek bir model olan BriFin'i sunuyoruz.

Anahtar sözcükler: Multipl Skleroz, Demiyelinizasyon, Tam Sayılı Programlama, Ağ Analizi, Protein-protein Etkileşimi.

Acknowledgement

I want to express my deepest gratitude to my advisor Associate Professor Can Alkan who first of all gave me the opportunity to work on such an exciting and motivating bioinformatics problem, answered my limitless questions, renewed my faith in conscientious science, and always supported me and believed in me. I am extremely grateful to him because of his supportive, constructive and conscientious attitude during the time we worked together, and for making it possible for my dream of contributing to science, which I had in Ankara Science High School, to come true.

I am also extremely grateful to Dr. Emre Karakoç for his guidance about the mathematical aspect of the problem, and to Associate Professor Bilal Ersen Kerman for his guidance regarding the biological aspect of the problem, and his eagerness for interdisciplinary collaboration. I would like to express my special thanks to Burcu Kurt Vatandaşlar for her great collaboration, our multidisciplinary consultations, which were very meaningful for me, and the experiments she conducted for this study.

I am also thankful to Dr. Ilknur Aydın Cantürk, M.D., Dr. Fatmagül İlayda Aydınlı, and Associate Professor Özge Arıcı Düz, M.D., for their contributions on the collection of the patient data to evaluate our study.

Thanks should also go to my colleagues Zülal, Sami and Ricardo for our valuable scientific discussions on the problem.

I would like to extend my thanks to the Scientific and Technological Research Council of Turkey (TÜBİTAK) for their support to me under the 2210-A program, support to Burcu Kurt Vatandaşlar under the 2211-C program, and further for their TÜBİTAK grant (1001-218S495) to Bilal Ersen Kerman.

Last but not the least, I would like to express my deepest gratitude to my lovely family for their invaluable efforts for me and for the environment convenient to wondering and questioning they gave me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Development of the bioinformatics field made it possible to analyze data in large quantities. DNA and protein sequencing data, protein interactions, gene-protein relations are instances of the generated data. By usage of biomedical technologies and mathematics, researchers revealed many meaningful results about biological processes, which are powerful supports or starting points for biological research. Most importantly, this research revealed useful information about disease mechanisms and disease-related biological actors from a holistic and data-driven view. There are many analyses conducted such as analysis of expressions of genes and proteins in patients, analysis of cell-level mechanisms for certain biological processes or analyses of genetic materials of organisms that were not not known before. This is a multidisciplinary field where the relevant research is conducted by diverse disciplines such as materials science, statistics, computer science, molecular biology and genetics [1].

When a disease is under investigation, at first, relevant biological data is analyzed. After understanding relations and revealing important information about the biological system in question, researchers hypothesize according to the found information, and do many experiments to test the searched hypotheses. Subsequently, this research reveals some information about the disease mechanisms.

When this information is enough to develop a treatment for the disease, researchers start to study on medicines that can cure the disease. In this step, they conduct various chemical and biological research. Consequently, a drug is developed after years of hard work to cure or treat the disease.

1.1 Autoimmune Diseases

Our immune system is very important for our health since it protects us against harmful organisms. When the immune system of the body overworks, it cannot recognize the body's own tissues or cells, and starts attacking them as if they are invaders. In autoimmune diseases the immune system attack the body's own cells, and today the reason of it is not exactly known.

There are various autoimmune diseases that affect different parts of the body such as Addison's disease, celiac, Hashimoto's thyroiditis, multiple sclerosis and Graves' disease [2]. Type 1 diabetes, multiple sclerosis, lupus, Crohn's disease, psoriasis, rheumatoid arthritis and scleroderma are among the most common autoimmune diseases [3].

1.2 A Common Autoimmune Disease: Multiple Sclerosis

Oligodendrocytes are specialized glial cells of the central nervous system (CNS) that surround axons with their plasma membrane and form the myelin sheath required for the proper functioning of the vertebrate nervous system. Owing to the insulation supplied by myelin sheath, action potentials propagate faster along myelinated axons compared to non-myelinated axons [4]. Furthermore, oligodendrocytes promote long-term neuronal survival and function [5, 6]. Loss of myelin (demyelination) causes the disruption of electrical signal transmission and exiguity of oligodendroglial support, which means damage to neurons. Demyelination

exists in several neurological diseases, and multiple sclerosis (MS) is one of the diseases where demyelination is most common [7, 8].

MS is a chronic, inflammatory, and neurodegenerative disease that affects over 2 million people and is characterized by demyelination that is observed in the various regions of the brain and spinal cord [9, 10, 11]. The etiology of the disease is thought to be immune disorder that arises from the interactions of genetic and environmental factors [12]. More particularly, abnormally activated immune cells attack oligodendrocytes and myelin, and damage them. Because of the dysregulation of interneuronal communication and signal transmission, several physical and cognitive symptoms are seen in MS patients. Currently, there is not a certain cure to treat MS completely, however there are some protective and preventive treatments that aim to prevent the attacks or decrease the frequency of the attacks and severity of the effects of them. Figure 1.1 illustrates myelin and demyelination.

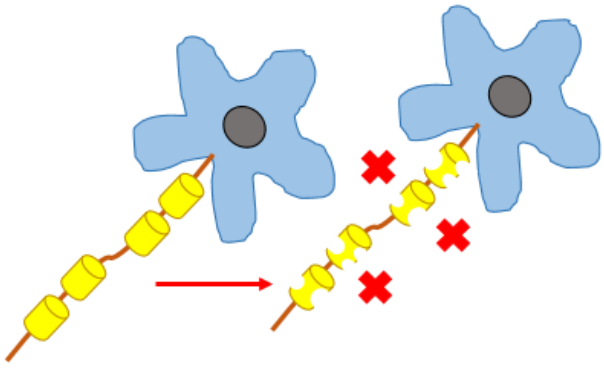

Figure 1.1: Demyelination. Represented in yellow, a healthy and damaged myelin sheath by immune cells.

The majority of patients show a relapsing-remitting phenotype of MS (RRMS) that is characterized by incidents of reversible neurological attacks that are just after followed by complete or local recovery. In time, neurological problems of patients diagnosed with RRMS become permanent, and secondary progressive MS (SPMS) develops [13, 14].

MS lesions are characterized by inflammation, axonal loss, and oligodendrocyte and neuronal cell death. The main immune cell types that contribute to inflammation in the lesions are macrophage, microglia and lymphocytes [11]. Therefore, it is important to analyze cell-to-cell interactions between the immune cells and the oligodendrocytes to reveal the cell-to-cell mechanisms behind MS.

1.3 Problem and Motivation

The fact that there is still no cure for multiple sclerosis manifests the need for further research to understand the disease mechanisms and identify the key actors in the disease. Identification of the key biological actors (i.e., proteins, genes, interactions) serves the understanding of the disease mechanisms, and consequently serves the development of the medicine that can cure the disease. To this end, as the first step of this complex research, our aim is to identify key actors of the disease from a network science perspective since it enables us to holistically analyze these complex systems (i.e., protein-protein interaction networks).

Network science has become an important discipline to analyze disease etiology in the last decade [15]. It is a powerful field for disease analysis since it evaluates the biological networks with a holistic and data-driven view instead of focusing on a single gene or mechanism. Diverse biological networks are analyzed for different purposes. In this study, we focused on the regular protein-protein interaction (PPI) networks formed between two cells, the target and the perpetrators of demyelination, since there is a problem with the proteomes of the cells involved in demyelination, resulting in a problem in cell-cell recognition.

For a PPI network, the question to which network medicine seeks an answer is whether the disease-associated proteins carry special attributes on the network or not. In this direction, we investigated "the most important" proteins, that are in fact specialized hubs, in the relevant PPI networks. Our motivation is that a change or perturbation in the most important nodes of the PPI network can cause bigger damage to the system due to the network properties of these

proteins, therefore these proteins have the potential to be the key disease proteins. Furthermore, our aim is to prioritize demyelination-related proteins, and present them for further biological research on demyelination and MS. Figure 1.2 shows the complex nature of a PPI network with an example, and emphasizes the importance of prioritization.

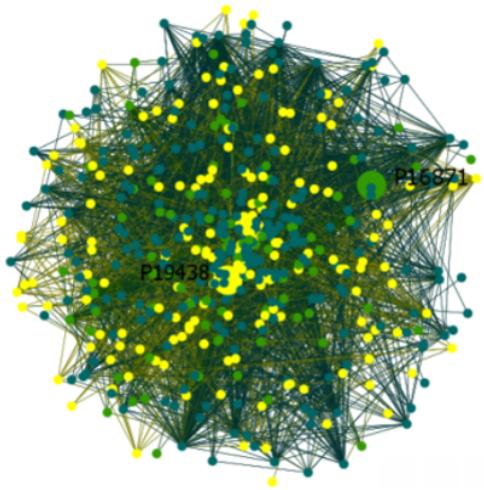

Figure 1.2: A PPI network example showing the complicated nature of PPI networks with thousands of nodes and edges.

1.4 Literature Review

Before the completion of the human genome project [16], there were candidate gene studies for MS focusing on a particular locus without properly knowing the population structure. With relatively limited data, these studies failed to detect robust genetic associations except the classic associations related to the major histocompatibility complex (MHC) region [17, 18]. After the completion of the human genome project, genetic variants started to be evaluated in a cost-effective and high throughput manner. This allowed the profiling of thousands of samples from many irrelevant cases and controls, and yielded comprehensive population-based association studies [17]. These genome-wide association studies (GWAS) with robust statistical adjustments unveiled many disease associations (i.e., more than 200 risk loci) [17]. Integration of a systems biology approach to

GWAS, strengthened the power of the research and increased precision, accuracy, confidence and coverage [19].

Another important technology commonly used in MS studies is expression microarray analysis. Expression microarrays enable the detection of expression changes between the MS cases and controls, which provides significant information regarding the disease genes/mechanisms especially when combined with a systems biology approach. The use of expression data in a gene/protein network enabled the identification of many novel disease proteins/genes [19]. For instance, Shang et al. [20] identified some potential key genes for MS by analyzing a protein–protein interaction (PPI) network and a gene co-expression regulatory network constructed with MS-related microRNAs (miRNAs). Xu et al. conducted another similar study, and they identified some key genes and miRNAs by using a protein-protein interaction network and a miRNA-mRNA regulatory network [21]. We can conclude that developing genetic technologies and a systems biology perspective enabled the discovery of many MS-related proteins/genes.

As for the studies that aim to identify novel disease genes/proteins based on network-based approaches, there are various methods in the literature whose number is rapidly increasing. One of the current research directions is to predict new disease-associated proteins using the existing knowledge on disease-associated proteins besides statistical and/or machine learning-based algorithms [22]. There are many research projects that aim to find new disease proteins/genes starting from the known disease proteins/genes on the network and investigating the direct neighbors or the nodes that are visited by a random walk. Several studies try to find associations between genes or proteins and diseases through genomic and transcriptomic data that is included in a gene co-expression or a protein interaction network, and previously identified disease-associated genes or proteins. For example, HIT'nDRIVE, proposed by Shrestha et al. [23] is a comprehensive computational method that integrates transcriptomic and genomic data and it aims to find the smallest set of patient-specific altered genes on the network that can cause transcriptional perturbations.

An alternative network analysis strategy is detecting hubs, which are nodes

that are few in number, yet have the highest degree in the network. Hub proteins are not necessarily disease-associated proteins; however, the reason why they should be investigated is the fact that they are involved in many interactions in the network [15]. Therefore, a change in their interactions results in relatively more significant biological alterations. Shang et al. [20] and Xu et al. [21] also utilized hubs in their research.

In this study, we identified the crucial proteins for the demyelination network by detecting hubs in the PPI networks formed within and between the cells that play a role in demyelination for further biological research. Here the hubs we refer to are "specialized" hubs, which connect two cells and we call "bridges" in the remainder of this thesis. Since we analyzed a network by combining two different networks through intercellular protein-protein interactions, we identified these bridges as the protein pairs that had the highest intracellular importance scores (see Methods), and that are involved in the highest number of the intercellular interactions. In this study, our aim is to detect bridges on the demyelination PPI network that might likely play a role in the development of MS. Please note that our bridge definition has no relation with the bridge definition in graph theory.

In contrast to analyzing a PPI network generated from healthy cells, analyzing the PPI network generated from disease-carrying cells is also a promising way to understand disease etiology. For example, Yurduseven et al. [24] identified MS biomarkers by performing interactome analysis using an MS-specific brain PPI network that was constructed using transcriptome data. Yurduseven et al. analyzed cell-type specific and cell-to-cell bridges by considering the degrees of the proteins on the MS-specific brain PPI networks.

Here we introduce BriFin (Bridge Finder) to detect the bridges in a cell-to-cell protein interaction network that consists of both inter- and intracellular interactions between two cell types. Using BriFin on healthy PPI networks, we identify proteins that may be associated with MS, specifically through playing key roles in immune cell-oligodendrocyte communications.

Revealing the proteins that take part in the disease mechanism is not an easy

task due to the difficulty of obtaining biological samples from the brain and spinal cord where MS indications are observed. Therefore, we evaluated BriFin by investigating MS-associated potential biomarkers in readily available peripheral blood mononuclear cell (PBMC) samples, which may contribute to the understanding of the disease mechanism, as well as being used in the diagnosis and follow-up of the disease. Among the proteins with the highest scores, we selected four that are likely important in MS pathogenesis and verified the expression levels of the genes that code these proteins in PBMCs of MS patients using quantitative real-time PCR. We showed that the expression levels of two out of four genes that code the suspect proteins were significantly decreased in MS patients, which suggests disrupted downstream networks.

Chapter 2

Methods

2.1 Dataset and Network Construction

Recent advances in sequencing technologies and increasing biological data available in public databases enable us to better model and understand cell-to-cell interactions and protein networks [25]. In this study, we obtained the proteome data for the three different cell types from several articles [26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37], and publicly available databases (UniProt [38] and The Human Protein Atlas [39]). Next, we downloaded the PPI network data for the intracellular interactions from the IntAct Molecular Interaction Database [40]. We identified the probable contact proteins (membrane and secreted proteins) and the intercellular interactions for each cell using the data of ligand-receptor pairs from the cell-to-cell communication databases (CellTalkDB [41] and Bader-Lab [42]). We then removed the unconnected nodes from the network and finally calculated PageRank and Betweenness Centrality scores for all nodes (i.e., proteins) using the Gephi network analysis tool [43]. In Figure 2.1, we present a visualization of the constructed network and the problem terminology, and in Figure 2.2 illustrations of the PPI networks of the cells we anaylzed.

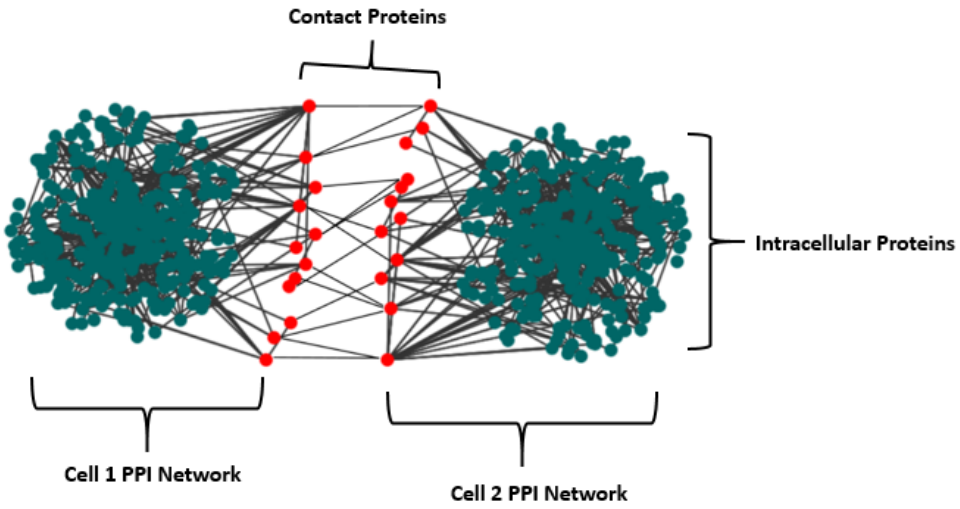

Figure 2.1: Constructed network and problem terminology

Figure 2.2: Illustrations of PPI Networks of the Analyzed Cells

2.2 Approach 1

We initially compiled data for two large interacting PPI networks, namely oligodendrocyte-macrophage and oligodendrocyte-T-cell networks, where all intra- and intercellular interactions are included. However, both of these PPI networks were very large: the oligodendrocyte-macrophage network contained

3,574 nodes and 37,314 edges, while the oligodendrocyte-T-cell network contained 11,813 nodes and 171,421 edges. While solving this network optimization problem, to handle the network of two cells as a whole, we formulated Formulation 1, where intracellular importance of a contact protein is based on the number of intracellular interactions that can be covered by selecting that protein. The model is based on the selection of contact protein pairs that can cover the specified amount of the intercellular and intracellular interactions. There are six decision variables: x_{ij} denoting whether the contact protein pair that consists of contact protein i from cell 1 and j from cell 2 is selected, y_{ij} denoting whether the interaction between the contact protein pair i, j is covered by the selected protein pairs, q_k denoting whether protein k of cell 1 is covered by the selected protein pairs, w_m denoting whether protein m of cell 2 is covered by the selected protein pairs, z_{kl} denoting whether the direct interaction between protein k and protein l of cell 1 is covered by the selected protein pairs, and u_{mn} denoting whether the direct interaction between protein m and protein n of cell 2 is covered by the selected protein pairs. Formulation 1 is as follows:

Formulation 1

$$
\begin{array}{lll}
\text{minimize} & \sum_{i\in A_1}\sum_{j\in B_1} x_{ij} & \\
\text{subject to} & (x_{il}+x_{tj})a_{ij} \le 2y_{ij}, & \forall i,t\in A_1, j,l\in B_1 \quad (1)\\
& y_{ij} \le (\sum_{l\in B_1} x_{il} + \sum_{t\in A_1} x_{tj})a_{ij}, & \forall i\in A_1, j\in B_1 \quad (2)\\
& \sum_{i\in A_1}\sum_{j\in B_1} y_{ij} \ge \alpha \sum_{i\in A_1}\sum_{j\in B_1} a_{ij} & (3)\\
& x_{ij} \le a_{ij}, & \forall i\in A_1, j\in B_1 \quad (4)\\
& y_{ij} \le q_i, & \forall i\in A_1, j\in B_1 \quad (5)\\
& y_{ij} \le w_j, & \forall i\in A_1, j\in B_1 \quad (6)\\
& q_i \le \sum_{j\in B_1} y_{ij}, & \forall i\in A_1 \quad (7)\\
& w_j \le \sum_{i\in A_1} y_{ij}, & \forall j\in B_1 \quad (8)\\
& q_k \le \sum_{i\in A_1} d_{ik}q_i, & \forall k\in A_2 \quad (9)\\
& d_{ik}q_i \le q_k, & \forall i\in A_1, k\in A_2 \quad (10)\\
& w_m \le \sum_{j\in B_1} e_{jm}w_j, & \forall m\in B_2 \quad (11)\\
& e_{jm}w_j \le w_m, & \forall j\in B_1, m\in B_2 \quad (12)\\
& (q_k+q_l)b_{kl} \le 2z_{kl}, & \forall k,l\in C \quad (13)\\
& z_{kl} \le (q_k+q_l)b_{kl}, & \forall k,l\in C \quad (14)\\
& (w_m+w_n)c_{mn} \le 2u_{mn}, & \forall m,n\in D \quad (15)\\
& u_{mn} \le (w_m+w_n)c_{mn}, & \forall m,n\in D \quad (16)\\
& \sum_{k\in C}\sum_{l\in C} z_{kl} \ge \beta_1 \sum_{k\in C}\sum_{l\in C} b_{kl}, & (17)\\
& \sum_{m\in D}\sum_{n\in D} u_{mn} \ge \beta_2 \sum_{m\in D}\sum_{n\in D} c_{mn}, & (18)
\end{array}
$$

In Formulation 1, A_1 is the set of contact proteins of cell 1, and A_2 is the non-contact (intracellular) proteins of cell 1, while B_1 is the set of contact proteins for cell 2, and B_2 is the set of intracellular proteins for cell 2. C denotes $A_1 \cup A_2$, and similarly D denotes $B_1 \cup B_2$.

a_{ij} is the binary parameter that denotes whether there is an interaction between contact protein i of cell 1 and contact protein j of cell 2. b_{kl} denotes whether there is a direct interaction between protein k and protein l of cell 1. c_{mn} denotes whether there is a direct interaction between protein m and protein n of cell 2. d_{ik} denotes whether there is a direct or indirect interaction between contact protein i and intracellular protein k of cell 1. Similarly, e_{jm} denotes whether there is a direct or indirect interaction between contact protein j and intracellular protein m of cell 2. Finally, α, β_1 and β_2 are the threshold parameters for the interactions to be covered. α is the minimum desired coverage rate of the intercellular interactions. β parameters are the minimum desired coverage rates of the intracellular interactions for cell 1 and cell 2, respectively. The reason of the use of separate coverage rate parameters for each cell is the possible different structure of the two PPI networks.

Constraints 1-4 ensure that a contact protein pair can be selected only if they are connected, and an intercellular interaction can be covered if at least one of the proteins in it is selected. Constraints 5-8 ensure that a contact protein is covered if and only if it is involved in the covered pairs. Constraints 9-12 ensure that an intracellular protein is covered if and only if there is a covered membrane protein that is directly or indirectly connected to it. Constraints 13-16 ensure that an intracellular interaction is covered if at least one of its proteins is covered. Finally, constraints 17 and 18 ensure that the desired coverage rate for the intracellular interactions is satisfied for each cell.

As the first step, we tested Formulation 1 by reducing the size of the problem. We solved the problem by including intracellular proteins only at most two layers away from the contact proteins. Since after a certain layer each contact protein is connected to an intracellular protein because of the small world effect of PPI networks, we limited the layers by two. That is, we assumed that the number of the connections between a contact protein and the intracellular proteins in the first two layers, is correlated with the total number of proteins in this cell that can be covered when this contact protein is selected. However, even in this case, the solution time was very long due to the nature of the problem, which is NP-hard [44]. Therefore, we designed an alternative approach, in which we reduce

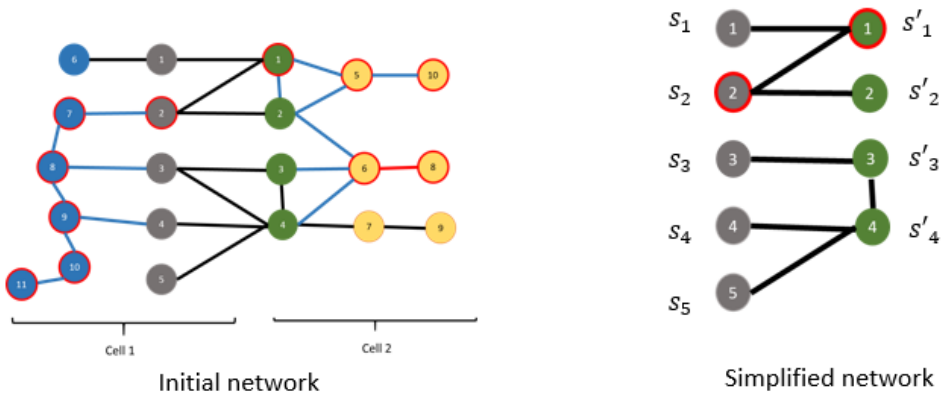

Figure 2.3: Representative Network Forms

the intracellular connections to importance scores for contact proteins, and we performed improvements on the second model.

In the alternative approach, we measured the importance of a contact protein in the relevant cell by taking the network centrality scores of the intracellular proteins and the distances of them to the contact protein into consideration. Although improvements in the design of the first formulation such as narrowing the search space by new kind of constraints or fewer number of decision variables are possible, we did not make improvements on this approach since they will not make a significant difference on the solution time due to the NP-hard nature of the problem. In addition, our second approach enabled us to make more diverse and comprehensive analyzes since it easily detects and presents network information and work much more faster. In Figure 2.3, we show exemplary network forms used for two approaches.

2.3 Approach 2

2.3.1 Assigning Intracellular Importance Scores to the Contact Proteins

As stated in the previous section, to reduce the problem size and to be able to solve it efficiently, we assigned an intracellular importance score (IIS) to each contact protein expressing its importance for the intracellular network of the cell it belongs to. To do this, we first determined each contact protein's direct and indirect interactors, and their distances using a breadth first search strategy. We evaluated the importance of the individual proteins in the intracellular network based on their PageRank centrality values since they express how likely these nodes are visited in the network. Combining this information, we assigned scores for each contact protein using Equation 1. Similarly, Ali Al-Fatlawi et al. [45] used PageRank metric in their network-based predictor NetRank to detect robust cancer biomarkers and showed their method had a strong prediction performance.

$$IIS_i = \sum_{j \in NC} \frac{a_{ij} PR_j}{d_{ij}} + \sum_{k \in C-i} \frac{a_{ik} PR_k}{d_{ik}}, \quad \forall i \in C \qquad \textbf{(Equation 1)}$$

In Equation 1, C denotes the set of contact proteins, and NC denotes the set of proteins that do not participate in intercellular interactions for the cell in question. PR_j denotes PageRank centrality score of protein j, a_{ij} denotes whether contact protein i and cell protein j are connected, while d_{ij} denotes the length of the shortest path between contact protein i and cell protein j in the network.

Since PPI networks have the small world phenomenon feature [46], which means a high connectivity between proteins causing each protein is connected to another at a certain layer, we developed a scoring method that can differentiate the scores of the contact proteins as much as possible. To this end, we calculated the IISs by evaluating the interactions between the contact proteins

separately and using a distance metric to measure the scores more sensitively. We used the network by removing the edges between the contact proteins to obtain the distance parameters in the first component of Equation 1, and we only used the network of the contact proteins to obtain the distance parameters in the second component. Our analyses prior to designing the formula also corrected the small world feature of the PPI networks.

There are diverse choices to assess the importance of a node of the network such as using one of the centrality metrics, using a combination of them or designing a new problem-specific metric. As previously mentioned, we chose to use PageRank centrality [47] to measure the importance of an individual node, since it reflects the level of involvement of a protein in intracellular interactions. PageRank algorithm assumes there is a random surfer on the network that follows the edges, and PageRank centrality of a node is based on the probability that the random surfer stops at that node. At any step, the probability that the person will continue the movement is called the damping factor. Damping factor used in this study was 0.85, which is the default parameter defined in Gephi. Also, the default value, which is 0.001, was used for parameter epsilon that is the convergence error threshold. Here, it is important to mention that this approach equally assesses intracellular interactions by not assigning particular importance to any interactions. Another centrality metric that we used in this study, Betweenness centrality [48], is based on the number of shortest paths that pass through that node. Although it is a commonly used metric in biological studies, we will be focusing on the results of the metric of our choice, PageRank, in the rest of the thesis, and will be also sharing the Betweenness-based results in Appendix.

2.3.2 Finding the Bridges

We define the most significant protein pairs (i.e., *bridges*) as the minimum number of contact protein pairs that can cover a certain majority (specified using a parameter named α) of the cell-to-cell interactions and have the highest IISs for the two intracellular networks. This problem is a variation of the set cover with

pairs problem [44]. The main difference is that our aim is not necessarily to cover all elements (edges) by minimum costly pairs of objects (nodes/proteins) in our problem, while in the set cover with pairs problem the goal is to cover all elements by the selected pairs. The reason we set α parameter instead of solving the set cover with pairs problem is that we aim to better prioritize the proteins and find a smaller set of proteins to be focused on in the further biological research. Therefore our formulation becomes an extension to the maximum quasi-clique problem [49], which asks to find a subgraph with the edge density of $\gamma \in (0,1)$ in a graph G=(V,E). In addition, we select the pairs among only the connected pairs instead of all possible pairs of nodes. This selection method provides biologically more meaningful protein pairs, since each selected pair is an intercellular interacting pair meaning that they connect networks of two cells, and they may be involved in demyelination.

Assigning scores to the contact proteins considerably reduces the problem size, and yields a bipartite network where only contact proteins exist with node scores. Using integer linear programming (ILP), we determine the bridges for the intercellular and intracellular networks that two cells form.

Our ILP model collectively evaluates the effect of the selected protein pairs on the network. So, it chooses the contact protein pairs that cover complementary intercellular interactions. It is also possible to evaluate the protein pairs separately. This can be done by sorting the nodes in descending order by the sum of the normalized degree of coverage (i.e., the number of edges a pair covers) and the normalized score of the pair. This sum is the overall importance score (OIS) of a protein pair. The score of a pair is calculated by summing the normalized IISs of the proteins in the pair.

2.3.2.1 ILP Model

We define P as the set of contact protein pairs (i.e., pairs of connected nodes) and E as the set of interactions (i.e., edges) in the network. Since we evaluate only the connected protein pairs, these two sets are equivalent in our implementation.

We further define c_{ij} as the binary parameter that indicates whether pair i covers interaction j. If one of the proteins in the pair is included in an interaction, it means that the pair covers that interaction. s_i is the parameter that indicates the inverse of the sum (i.e., 1/sum) of the normalized IISs of the proteins in pair i. Since this is a minimization problem where we want to find the minimum number of proteins with highest scores that cover the highest number of the edges, we assign a score to each pair which is equal to the inverse of the sum of the scores of the proteins in that pair. Finally, α parameter expresses the minimum desired coverage rate of the interactions in the network. When α equals to 1, the problem becomes finding protein pairs that cover all of the interactions in the network (the set cover with pairs problem). In the ILP model, we have two binary decision variables x and y. x_i expresses whether pair i is selected, while y_j expresses whether interaction j is covered by the selected protein pairs. Formulation 2 for the problem is as follows:

Formulation 2

minimize	$\sum_{i \in P} x_i s_i$		
subject to	$x_i c_{ij} \leq y_j,$	$\forall i \in P, j \in E$	(1)
	$\sum_{j \in E} y_j \geq \alpha \lvert E \rvert,$		(2)
	$y_j \leq \sum_{i \in P} x_i c_{ij},$	$\forall j \in E$	(3)

The first constraint ensures that an edge is covered if a pair covering it is selected. The second constraint provides that the number of the covered interactions satisfies the minimum desired coverage rate. Finally, the third constraint ensures an edge cannot be covered if a pair covering it is not selected. We solved the ILP using the Python Gurobi Solver [50].

2.4 Evaluating BriFin with MS patient data

To test the MS-association of the proteins in the bridges detected by BriFin, we collaborated with researchers from biology domain. By evaluating BriFin results and biological aspects, we chose 4 proteins to test. We conducted mRNA experiments on the samples collected from a group of MS patients. For the self-completeness of the thesis, we here present the details of the experiments.

2.4.1 Subjects and Isolation of Peripheral Mononuclear Cells (PBMCs)

We included eight Relapsing-Remitting MS (RRMS) patients and seven age and gender-matched healthy individuals as controls in this study. All patients are currently treated with immunomodulatory drugs. We presented the demographic and clinical features as well as immunomodulatory drugs of patients and healthy controls in Supplementary Table 1. After receiving written informed consent from all participants, we collected venous blood samples using tubes containing ethylene diamine tetra-acetic acid (EDTA). We separated PBMCs from venous blood via density gradient centrifugation using Lymphocyte Separation Medium, Density 1.077 g/ml (Capricorn). Finally, we pelleted the isolated PBMCs and stored them at -80° C for further use.

2.4.2 Total RNA isolation and Quantitative Real-Time PCR (qRT-PCR)

Total RNA was isolated from PBMCs with TriGent reagent (Biomatik) according to the manufacturer's instructions. Quantification of RNA samples was conducted using the Nanodrop Spectrophotometer (ThermoFisher Scientific). RNA samples were reverse transcribed using the Advanced cDNA Synthesis Kit (Wisent Bio-products) as recommended by the manufacturer. qRT-PCR was conducted on

BioRad CFX Connect using The SensiFAST SYBR No-ROX Kit (Meridian Bioscience) with the primer pairs of HSP90AA1, HSP90B1, CALR, and TFRC (For list of the primer sets, see Supplementary Table 2). The relative expression levels of each transcript were calculated by normalizing them against the expression of the housekeeping gene β-actin (ACTB). For the fold change analysis, transcript levels were compared to the control group.

2.4.3 Statistical analysis

We performed the statistical analysis of the results using GraphPad Prism version 8.0.0 (GraphPad Software, San Diego, California, USA). We used the non-parametric Mann-Whitney unpaired test to analyze the data, and we considered $p < 0.05$ to be statistically significant.

Chapter 3

Results and Discussion

We tested the performance of BriFin using PPI interaction networks, cell-to-cell communication databases and proteome data generated for oligodendrocytes, macrophages, and T-cells downloaded from different sources as explained in the Methods section. The number of all proteins and contact proteins (those involved in intercellular interactions) in each cell type are given in Table 3.1.

Table 3.1: Proteome sizes of each cell type used in this study

Cell Type	**Number of contact proteins**	**Total number of proteins**
Oligodendrocyte	282	2,846
T-cell	647	8,967
Macrophage	210	728

After constructing the PPI networks summarized in Table 1, we calculated IISs of the contact proteins. We reduced the intracellular network connections to these scores to be able to solve the problem efficiently. This resulted in a bipartite network with node scores where we investigate the bridges among the protein pairs, each of which consists of one contact protein from cell 1 and one contact protein from cell 2. We solved the ILP model for 2,191 contact protein pairs for the oligodendrocyte-macrophage network, and 5,983 contact protein pairs for the oligodendrocyte-T-cell network.

To prioritize the contact protein pairs, we used different thresholds for α, which

denotes the minimum ratio of cell-to-cell interactions covered by the selected contact protein pairs. Table 3.2 presents the number of the selected protein pairs for each setting of α for the two networks we analyzed in this study.

Table 3.2: Number of the selected bridges by BriFin for different α values for the two networks

Network / α	0.1	0.2	0.4	0.6	0.8
Oligodendrocyte-macrophage	2	5	11	21	36
Oligodendrocyte-T-cell	3	7	18	35	62

We identified that some of the selected proteins by the model have been associated with MS, other autoimmune diseases and neurological diseases in the literature. Here, we present the results for the two smallest α values: 0.1 and 0.2, based on the PageRank scores, and the relevant MS-associations. For the oligodendrocyte-macrophage network, two bridges were selected for α=0.1 by BriFin. These bridges are ITGB1 (Integrin Beta-1) - HSP90AA1 (Heat Shock Protein 90-alpha) and HSP90AA1 - TFRC (Transferrin Receptor Protein 1), where the first protein belongs to oligodendrocyte and the second to macrophage. In the relevant literature, variants of ITGB1 are found being associated with MS [51]. In a study where MS-specific membrane-associated biomarkers were investigated in experimental autoimmune encephalomyelitis (EAE), an animal model of MS, it was shown that TFRC protein expression was down-regulated in PBMCs [52]. In our study, we experimentally tested TFRC in blood samples of MS patients, and found that its mRNA expression is decreased in MS patients.

For the same network, 5 bridges were selected by BriFin for α=0.2 which are HMGB1 (High Mobility Group Box 1) - VCAM1 (Vascular Cell Adhesion Molecule 1), HSP90AA1 - TFRC, ITGB1 - ANXA2 (Annexin A2), FN1 (Fibronectin 1) - PKM (Pyruvate Kinase M1/2) and APP (Amyloid Beta Precursor Protein) - HSP90AA1. We see that some proteins are again selected for the higher α value in either the same or different bridges. To the best of our knowledge, all of these selected proteins are associated with MS in the literature except HSP90AA1 whose association we present in this study. It is also important to observe that HSP90AA1 is important for both the immune cells and the oligodendrocyte cell.

In Figure 3.1, illustrations of the selected bridges for α=0.1 and α=0.2 for the oligodendrocyte-macrophage network are shown.

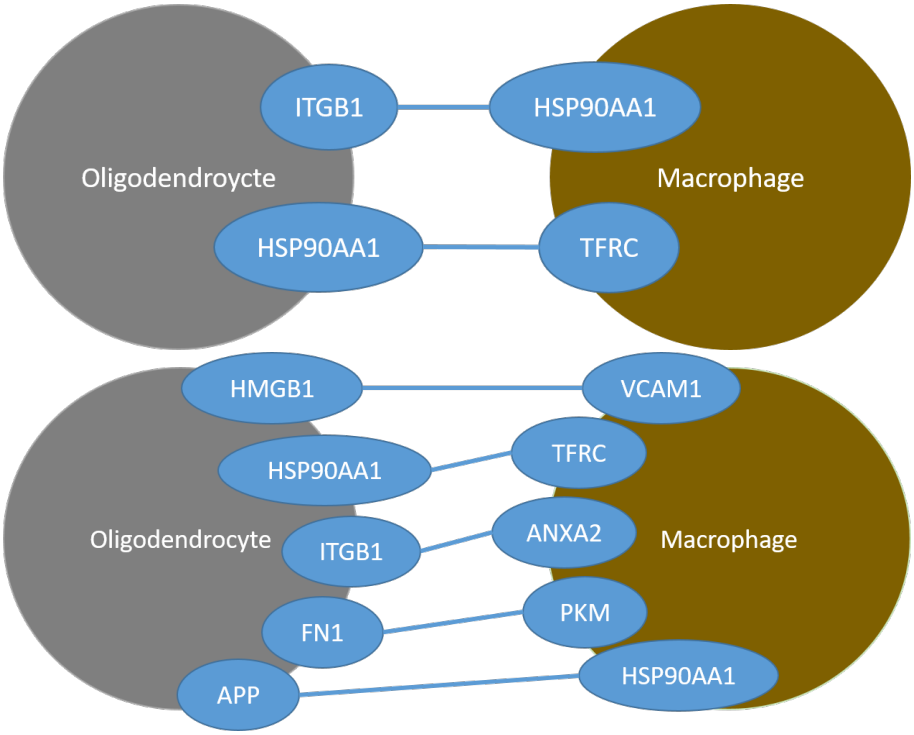

Figure 3.1: Selected bridges in the oligodendrocyte-macrophage network for α=0.1 and α=0.2 respectively

Regarding these proteins, it was shown that HMGB1 expression levels were increased in PBMCs of MS patients significantly [53, 54], and various VCAM1 positive microglia/macrophages exist at the edges of MS lesions [55]. Also, VCAM1 and its variants [51] are associated with MS in a considerable number of studies. Another study based on the EAE model [56] emphasizes the therapeutic potential of PKM2 activators in MS-like diseases, and shows how these activators change T-cell function. That APP plays a role in MS was shown in several studies [57, 58]. Gehrmann et al. [57] showed that the level of APP expression is correlated with histopathological lesion development, therefore APP is an important biomarker for the progression of MS. Also, Matías-Guiu et al. stated APP has a role in both demyelination and remyelination [58]. Variants of FN1 are shown to be associated with multiple sclerosis in the study of Dardiotis et al. [51]. Iparraguirre et al. reported that ANXA2 expression is a good biomarker for MS [59].

For the oligodendrocyte-T-cell network, 3 bridges are selected by BriFin for α=0.1. These are APP - APP, HSP90AA1 - EGFR (Epidermal Growth Factor Receptor) and FN1 - ITGB1 in the respective order of the cells. There are some proteins selected in common with the oligodendrocyte- macrophage network such as APP, HSP90AA1, FN1 and ITGB1. Scalabrino [60] stated that recent findings show that EGF expression is significantly decreased in the cerebrospinal fluid (CSF) and spinal cord (SC) of the MS patients, and the new information about the role of EGF in MS required a critical reassessment of the MS pathogenesis.

For the same network, 7 bridges are detected by BriFin for α=0.2: CLU (clusterin) - EGFR, PKM - ANXA2, ANXA2 - PKM, ITGB1 - FN1, HSP90AA1 - HSP90AA1, APP - APP and FN1-ITGB1. That the similar proteins are detected and there are bridges consisting of the same protein is worth mentioning. In Figure 3.2, illustrations of the selected bridges for α=0.1 and α=0.2 for the oligodendrocyte-T-cell network are shown.

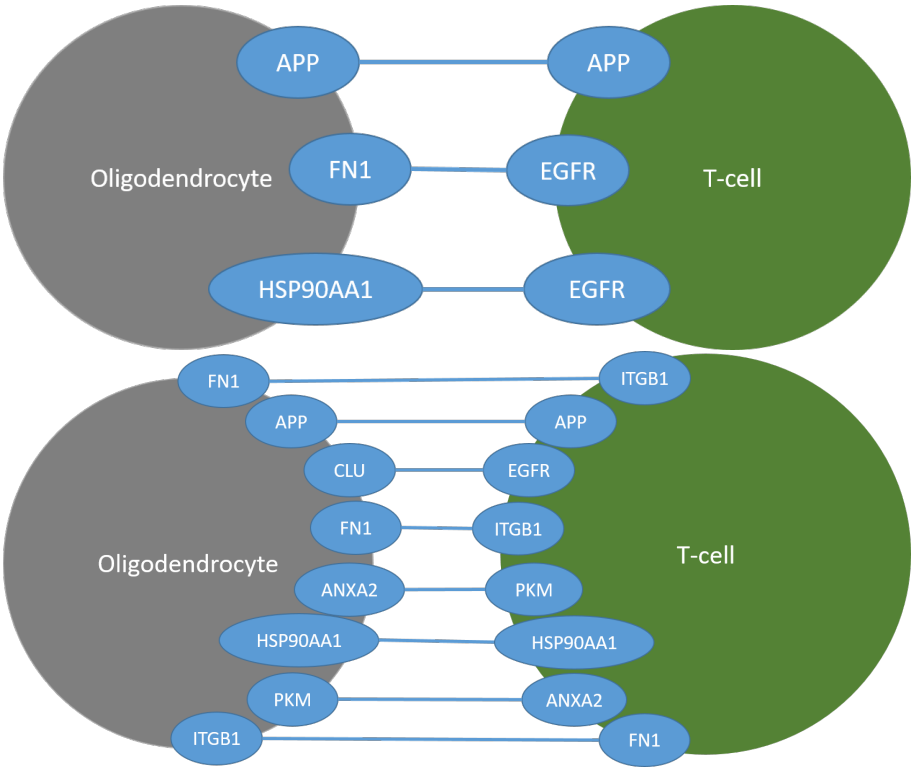

Figure 3.2: Selected bridges in the oligodendrocyte-T-cell network for α=0.1 and α=0.2 respectively

For larger α values, there are also MS-associated proteins such as CD44 (CD44 Antigen), APOE (Apolipoprotein E), ALDOA (Aldolase A), IL7R (Interleukin-7 Receptor) and HLA-DRB1 (Major Histocompatibility Complex, Class II, DR Beta 1) among the selected proteins by BriFin. In the domain of multiple sclerosis research, there are many studies about the role of IL7R for the disease. For instance, Lei et al. [61] showed that IL7R is down-regulated during demyelination, and by targeted knockdown, they showed that IL7R is crucial for myelination in embryonic and larval zebrafish in a zebrafish model. Barcellos et al. [62] reported that there is a strong association between certain variants of HLA-DRB1 and multiple sclerosis in a comprehensive study that includes data from diverse populations. Farias et al. [63] reported that ALDOA and APOE genes are upregulated in CSF of MS patients. In addition, Guan et al. [64] showed that CD44 controls the development of EAE.

Since selected bridges whose both proteins are MS-associated may provide useful information about the disease mechanism, here we mention some of these bridges (first protein belongs to oligodendrocyte, the second to immune cell): APOE - VCAM1, HLA-DRB1 - PKM, APOE - ALDOA, APOE - ANXA2, FN1 - CD44.

The full lists of the selected proteins for all α values are given in Tables A.3-A.12. We also report the selected protein pairs when the Betweenness Centrality metric is used to calculate IISs along with the OISs of the protein pairs since it is a commonly-used metric. This analysis yields similar results to the PageRank-based analysis with some changes in the prioritization order. Similarly, we give the full lists in Tables A.13-A.22. In Table 3.3, we show top 10 highest-scoring (in terms of OIS) bridges among the selected ones by BriFin for all tested α values.

It is important to note that, due to the protein pairs that have similar scores, many alternative optimal solutions exist for some α values. However, these solutions include mostly the same proteins. In addition, the model evaluates the collective effect of the protein pairs on the network, which means it chooses the highest scoring protein pairs whose interactions are complementary. Therefore, it is different than selecting the individually highest scoring pairs. Protein pairs

Table 3.3: Top 10 bridges selected by BriFin for all tested α values sorted by highest OIS

α	**Oligodendrocyte-macrophage network** (proteins in respective order)	α	**Oligodendrocyte-T-cell network** (proteins in respective order)
0.2	APP-HSP90AA1	0.1	FN1-ITGB1
0.2	FN1-PKM	0.1	APP-APP
0.1	HSP90AA1-TFRC	0.4	FN1 ITGAV
0.1	ITGB1-HSP90AA1	0.1	HSP90AA1-EGFR
0.2	ITGB1-ANXA2	0.8	FN1-ITGA4
0.4	PKM-HSP90AA1	0.2	ITGB1-FN1
0.4	FN1-CD44	0.4	APP-TFRC
0.4	HSP90AA1-CTSD	0.2	HSP90AA1-HSP90AA1
0.4	APP-CALR	0.4	ITGB1-C1QBP
0.4	CTSD-PKM	0.2	CLU-EGFR

with low importance scores might be among the selected pairs if they cover some interactions that are not covered by the other proteins.

We list the individual scores for the top 40 protein pairs (OISs) in Tables A.23-A.26, and share the complete lists online. Under this scoring metric, the top 10 bridges for the two networks are shown in Table 3.4 (with PageRank based scores).

Table 3.4: Top 10 bridges for each network based on OISs

Order	**Oligodendrocyte-macrophage network** (proteins in respective order)	**Oligodendrocyte-T-cell network** (proteins in respective order)
1	APP-PKM	FN1-EGFR
2	APP-HSP90AA1	FN1-APP
3	HSP90AA1-PKM	FN1-HSP90AA1
4	HSP90AA1-HSP90AA1	FN1-FN1
5	FN1-PKM	FN1-ITGB1
6	FN1-HSP90AA1	FN1-PKM
7	APP-TFRC	FN1-TFRC
8	HSP90AA1-TFRC	FN1-ANXA2
9	FN1-TFRC	APP-EGFR
10	APP-ANXA2	FN1-ALDOA

In our analysis, we also identified the top contributors to the scores of the selected contact proteins among the intracellular proteins to prioritize the intracellular proteins. Closer proteins on the network are likely to contribute more because of the formula we used to calculate IISs. However, individual PageRank

scores are also effective. That is, the top contributors for a contact protein may be interpreted as the ones having the highest centrality scores among the proteins close to it on the network. For each cell, we identified the top 20 score contributors for the selected highest scoring 10 proteins. We show the most frequent intracellular contributors among the top 20 in Table A.27, for each cell.

To experimentally validate our predictions, we selected 4 proteins among the highest scoring pairs that likely play a role in demyelination and investigated their mRNA expression levels in the blood samples obtained from MS patients. These proteins are HSP90AA1, HSP90B1, CALR, and TFRC. The expression levels of the selected proteins' transcripts in PBMCs were determined by qRT-PCR. Our results showed that the expression levels of the HSP90AA1, and TFRC reduced significantly in the MS group when compared to the control group (Figure 3.3).

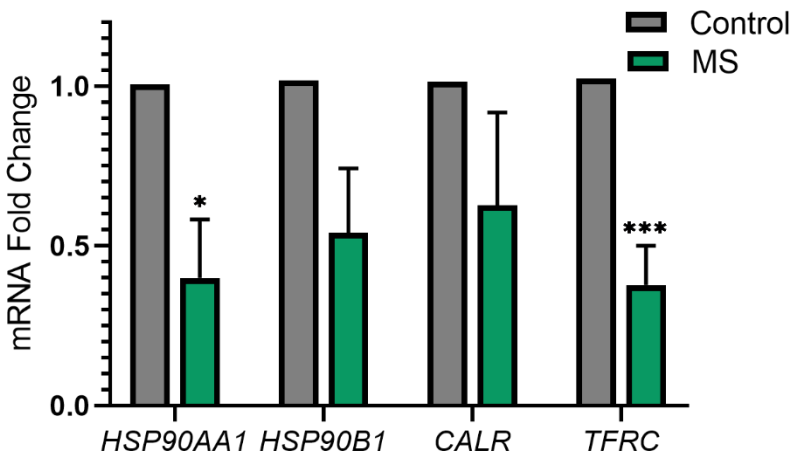

Figure 3.3: HSP90AA1, HSP90B1, CALR, and TFRC mRNA expressions in PBMCs of control and MS patients. Selected genes were analyzed by qRT-PCR. Fold changes in expression levels of MS patients (n=8, green bars) compared with the healthy controls (n=7, gray bars) were shown in the graph bar. qRT-PCR analysis indicates that the mRNA levels of the HSP90AA1 and TFRC were significantly low in the MS group relative to healthy controls. The error bars are presented as mean ± SEM, p values; * < 0.05, *** < 0.001, Mann Whitney U test.

To assess the performance of BriFin for larger α values, we further conducted a problem-specific literature search for the selected proteins by BriFin in the

oligodendrocyte-macrophage network, which is relatively smaller. We investigated the associations of each protein with MS, related to the cell it belongs to. When we could not find cell type-specific associations, we used other associations that may be related to the cell that the protein in question belongs to. Articles [51-91] include the information regarding these associations. In Table 3.5, the rate of the MS-associated proteins among the selected proteins by BriFin for all α values in the oligodendrocyte-macrophage network are shown.

Table 3.5: Rate of MS-associated proteins among the proteins selected by BriFin in the oligodendrocyte-macrophage network for different α values

α	Rate of MS-associated proteins
0.1	100%
0.2	100%
0.4	81%
0.6	80%
0.8	77%

For smaller α values, the rate of MS-associated proteins is higher, which supports our assumption that smaller alpha values detect more important proteins.

Chapter 4

Conclusion

Network science is an essential tool to infer physiological interpretation from biological networks since it evaluates the networks with a holistic view, and also a good way to support biological studies since the resources are limited and there is a great deal of relevant data to eliminate and prioritize. Here, we presented the BriFin model to detect bridges, key protein-protein pairs, between oligodendrocytes and macrophages or T cells. We showed that the detected proteins by our model were associated with MS, and two detected proteins were differentially expressed in MS patients in an application of network analysis. That the hubs detected by the model are also important proteins to investigate because of the biological mechanisms they are involved in is a meaningful result, and proteins/genes that are both biologically and mathematically pointed out might be good starting points to do more research on.

Among the selected protein pairs for all the tested alpha values, there are pairs whose both proteins are MS-associated. Investigating the biological mechanisms behind the interaction of the proteins in these pairs may yield useful information to understand MS better. In addition, proteins that are the matches of the proteins associated with multiple sclerosis are good research targets for further studies. Also, research on selected proteins that are associated with autoimmune diseases and other neurodegenerative diseases may yield useful information about

MS. Finally, the selected proteins and the highest-scoring pairs that have not been associated with any disease yet might be potential research directions.

The provided results are based on the collected interaction and proteome data. Therefore, the quality of the results depends on the quantity and the quality of the data. These computational results can become more reliable and quality by more data and more biological expertise. The distance of the biological assumptions from reality and the level of the inclusion of these assumptions are important for computational studies, and improvements on these topics might be new research directions. In addition, the difficulty of obtaining data from patients makes computational studies harder to evaluate, and the improvements about patient data accession can support the development of computational studies.

In the literature, there are many disease-associated proteins with different explanations. Each protein has a different meaning and importance for the cell. While some proteins are the actual causal proteins that originate the demyelination problem, other proteins are the ones that are affected by them. Therefore, the changes in them are the consequences of the changes in the actual disease-causing proteins. In consequence, elaborate studies are needed to discover the actual mechanisms, and along this discovery path, computational studies can provide meaningful contributions to the biological studies.

Our network analysis approach might be useful for other diseases where two cell types interact such as autoimmune diseases, cancer, many neurological diseases and for research areas in which cell-to-cell interactions are dominant such as immunotherapy and microbiome-host interaction.

Bibliography

[1] A. Bayat *et al.*, "Science, medicine, and the future: Bioinformatics," *British Medical Journal*, vol. 324, pp. 1018–1022, 2002.

[2] `https://my.clevelandclinic.org/health/diseases/21624-autoimmune-diseases`. Accessed: 2022-12-26.

[3] `https://medlineplus.gov/ency/article/000816.htm`. Accessed: 2022-12-26.

[4] D. Hartline and D. Colman, "Rapid conduction and the evolution of giant axons and myelinated fibers," *Current Biology*, vol. 17, no. 1, pp. R29–R35, 2007.

[5] A. S. Saab and K.-A. Nave, "Myelin dynamics: protecting and shaping neuronal functions," *Current opinion in neurobiology*, vol. 47, pp. 104–112, 2017.

[6] U. Fünfschilling, L. M. Supplie, D. Mahad, S. Boretius, A. S. Saab, J. Edgar, B. G. Brinkmann, C. M. Kassmann, I. D. Tzvetanova, W. Möbius, *et al.*, "Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity," *Nature*, vol. 485, no. 7399, pp. 517–521, 2012.

[7] M. Traka, J. R. Podojil, D. P. McCarthy, S. D. Miller, and B. Popko, "Oligodendrocyte death results in immune-mediated cns demyelination," *Nature neuroscience*, vol. 19, no. 1, pp. 65–74, 2016.

[8] M. Kipp, "Oligodendrocyte physiology and pathology function," 2020.

[9] C. Walton, R. King, L. Rechtman, W. Kaye, E. Leray, R. A. Marrie, N. Robertson, N. La Rocca, B. Uitdehaag, I. van der Mei, *et al.*, "Rising

prevalence of multiple sclerosis worldwide: Insights from the atlas of ms," *Multiple Sclerosis Journal*, vol. 26, no. 14, pp. 1816–1821, 2020.

[10] C. Stadelmann, S. Timmler, A. Barrantes-Freer, and M. Simons, "Myelin in the central nervous system: structure, function, and pathology," *Physiological reviews*, vol. 99, no. 3, pp. 1381–1431, 2019.

[11] N. Garg and T. W. Smith, "An update on immunopathogenesis, diagnosis, and treatment of multiple sclerosis," *Brain and behavior*, vol. 5, no. 9, p. e00362, 2015.

[12] T. Olsson, L. F. Barcellos, and L. Alfredsson, "Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis," *Nature Reviews Neurology*, vol. 13, no. 1, pp. 25–36, 2017.

[13] C. Lucchinetti, R. Gavrilova, I. Metz, J. Parisi, B. Scheithauer, S. Weigand, K. Thomsen, J. Mandrekar, A. Altintas, B. Erickson, *et al.*, "Clinical and radiographic spectrum of pathologically confirmed tumefactive multiple sclerosis," *Brain*, vol. 131, no. 7, pp. 1759–1775, 2008.

[14] R. Rajendran, G. Böttiger, C. Stadelmann, S. Karnati, and M. Berghoff, "Fgf/fgfr pathways in multiple sclerosis and in its disease models," *Cells*, vol. 10, no. 4, p. 884, 2021.

[15] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease," *Nature reviews genetics*, vol. 12, no. 1, pp. 56–68, 2011.

[16] I. H. G. S. Consortium, "Finishing the euchromatic sequence of the human genome," *Nature*, vol. 431, no. 7011, pp. 931–945, 2004.

[17] C. Cotsapas and M. Mitrovic, "Genome-wide association studies of multiple sclerosis," *Clinical & translational immunology*, vol. 7, no. 6, p. e1018, 2018.

[18] N. A. Patsopoulos, "Genetics of multiple sclerosis: an overview and new directions," *Cold Spring Harbor perspectives in medicine*, vol. 8, no. 7, p. a028951, 2018.

[19] I. Diakou, E. Papakonstantinou, L. Papageorgiou, K. Pierouli, K. Dragoumani, D. A. Spandidos, F. Bacopoulou, G. P. Chrousos, G. N. Goulielmos, E. Eliopoulos, *et al.*, "Multiple sclerosis and computational biology," *Biomedical Reports*, vol. 17, no. 6, pp. 1–7, 2022.

[20] Z. Shang, W. Sun, M. Zhang, L. Xu, X. Jia, R. Zhang, and S. Fu, "Identification of key genes associated with multiple sclerosis based on gene expression data from peripheral blood mononuclear cells," *PeerJ*, vol. 8, p. e8357, 2020.

[21] Z.-b. Xu, X. Feng, W.-n. Zhu, M.-l. Qiu, *et al.*, "Identification of key genes and micrornas for multiple sclerosis using bioinformatics analysis," *Medicine*, vol. 100, no. 48, pp. e27667–e27667, 2021.

[22] P. Tieri, L. Farina, M. Petti, L. Astolfi, P. Paci, and F. Castiglione, "Network inference and reconstruction in bioinformatics," 2019.

[23] R. Shrestha, E. Hodzic, T. Sauerwald, P. Dao, K. Wang, J. Yeung, S. Anderson, F. Vandin, G. Haffari, C. C. Collins, *et al.*, "Hit'ndrive: patient-specific multidriver gene prioritization for precision oncology," *Genome research*, vol. 27, no. 9, pp. 1573–1588, 2017.

[24] K. Yurduseven, Y. K. Babal, E. Celik, B. E. Kerman, and I. A. Kurnaz, "Multiple sclerosis biomarker candidates revealed by cell-type-specific interactome analysis," *OMICS: A Journal of Integrative Biology*, vol. 26, no. 5, pp. 305–317, 2022.

[25] E. Armingol, A. Officer, O. Harismendy, and N. E. Lewis, "Deciphering cell–cell interactions and communication from gene expression," *Nature Reviews Genetics*, vol. 22, no. 2, pp. 71–88, 2021.

[26] K. Iwata, C. C. Café-Mendes, A. Schmitt, J. Steiner, T. Manabe, H. Matsuzaki, P. Falkai, C. W. Turck, and D. Martins-de Souza, "The human oligodendrocyte proteome," *Proteomics*, vol. 13, no. 23-24, pp. 3548–3553, 2013.

[27] A. Ishii, R. Dutta, G. M. Wark, S.-I. Hwang, D. K. Han, B. D. Trapp, S. E. Pfeiffer, and R. Bansal, "Human myelin proteome and comparative analysis with mouse myelin," *Proceedings of the National Academy of Sciences*, vol. 106, no. 34, pp. 14605–14610, 2009.

[28] P. de Monasterio-Schrader, O. Jahn, S. Tenzer, S. P. Wichert, J. Patzig, and H. B. Werner, "Systematic approaches to central nervous system myelin," *Cellular and Molecular Life Sciences*, vol. 69, no. 17, pp. 2879–2894, 2012.

[29] R. N. Joshi, C. Stadler, R. Lehmann, J. Lehtiö, J. Tegnér, A. Schmidt, and M. Vesterlund, "Tcellsubc: an atlas of the subcellular proteome of human t cells," *Frontiers in immunology*, vol. 10, p. 2708, 2019.

[30] M. Pagani, M. Rockstroh, M. Schuster, G. Rossetti, M. Moro, M. Crosti, and J. M. Tomm, "Reference proteome of highly purified human th1 cells reveals strong effects on metabolism and protein ubiquitination upon differentiation," *Proteomics*, vol. 15, no. 21, pp. 3644–3647, 2015.

[31] A. Dupont, C. Tokarski, O. Dekeyzer, A.-L. Guihot, P. Amouyel, C. Rolando, and F. Pinet, "Two-dimensional maps and databases of the human macrophage proteome and secretome," *Proteomics*, vol. 4, no. 6, pp. 1761–1778, 2004.

[32] M.-C. Slomianny, A. Dupont, F. Bouanou, O. Beseme, A.-L. Guihot, P. Amouyel, J.-C. Michalski, and F. Pinet, "Profiling of membrane proteins from human macrophages: comparison of two approaches," *Proteomics*, vol. 6, no. 8, pp. 2365–2375, 2006.

[33] S. Eligini, M. Brioschi, S. Fiorelli, E. Tremoli, S. Colli, and C. Banfi, "Data for proteomic analysis of human monocyte-derived macrophages," *Data in Brief*, vol. 4, pp. 177–179, 2015.

[34] R. A. S. Raposo, B. Thomas, G. Ridlova, and W. James, "Proteomic-based identification of cd4-interacting proteins in human primary macrophages," *PLoS One*, vol. 6, no. 4, p. e18690, 2011.

[35] R. Lichtenfels, G. Rappl, A. A. Hombach, C. V. Recktenwald, S. P. Dressler, H. Abken, and B. Seliger, "A proteomic view at t cell costimulation," *PloS one*, vol. 7, no. 4, p. e32994, 2012.

[36] C. J. Mitchell, D. Getnet, M.-S. Kim, S. S. Manda, P. Kumar, T.-C. Huang, S. M. Pinto, R. S. Nirujogi, M. Iwasaki, P. G. Shaw, *et al.*, "A multi-omic

analysis of human naïve cd4+ t cells," *BMC systems biology*, vol. 9, no. 1, pp. 1–16, 2015.

[37] A. Graessel, S. M. Hauck, C. von Toerne, E. Kloppmann, T. Goldberg, H. Koppensteiner, M. Schindler, B. Knapp, L. Krause, K. Dietz, *et al.*, "A combined omics approach to generate the surface atlas of human naive cd4+ t cells during early t-cell receptor activation," *Molecular & Cellular Proteomics*, vol. 14, no. 8, pp. 2085–2102, 2015.

[38] A. Bairoch, R. Apweiler, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, *et al.*, "The universal protein resource (uniprot)," *Nucleic acids research*, vol. 33, no. suppl_1, pp. D154–D159, 2005.

[39] M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, *et al.*, "Proteomics. tissue-based map of the human proteome.," *Science (New York, NY)*, vol. 347, no. 6220, pp. 1260419–1260419, 2015.

[40] H. Hermjakob, L. Montecchi-Palazzi, C. Lewington, S. Mudali, S. Kerrien, S. Orchard, M. Vingron, B. Roechert, P. Roepstorff, A. Valencia, *et al.*, "Intact: an open source molecular interaction database," *Nucleic acids research*, vol. 32, no. suppl_1, pp. D452–D455, 2004.

[41] X. Shao, J. Liao, C. Li, X. Lu, J. Cheng, and X. Fan, "Celltalkdb: a manually curated database of ligand–receptor interactions in humans and mice," *Briefings in bioinformatics*, vol. 22, no. 4, p. bbaa269, 2021.

[42] "Baderlab cell-cell interaction database." `https://baderlab.org/CellCellInteractions`. Accessed: 2021-11-02.

[43] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks," in *Proceedings of the international AAAI conference on web and social media*, vol. 3, pp. 361–362, 2009.

[44] R. Hassin and D. Segev, "The set cover with pairs problem," in *International Conference on Foundations of Software Technology and Theoretical Computer Science*, pp. 164–176, Springer, 2005.

[45] A. Al-Fatlawi, N. Afrin, C. Ozen, N. Malekian, and M. Schroeder, "Netrank recovers known cancer hallmark genes as universal biomarker signature for cancer outcome prediction," *Frontiers in bioinformatics*, vol. 2, 2022.

[46] https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/protein-protein-interaction-networks/properties-of-ppins-small-world-effect/

[47] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web.," tech. rep., Stanford InfoLab, 1999.

[48] https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/building-and-analysing-ppins/topological-ppin-analysis/betweenness-centrality/.

[49] J. Pattillo, A. Veremyev, S. Butenko, and V. Boginski, "On the maximum quasi-clique problem," *Discrete Applied Mathematics*, vol. 161, no. 1-2, pp. 244–257, 2013.

[50] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2022.

[51] E. Dardiotis, E. Panayiotou, V. Siokas, A.-M. Aloizou, K. Christodoulou, A. Hadjisavvas, M. Pantzaris, N. Grigoriadis, G. M. Hadjigeorgiou, and T. Kyriakides, "Gene variants of adhesion molecules predispose to ms: A case-control study," *Neurology Genetics*, vol. 5, no. 1, 2019.

[52] L. F. Dagley, N. P. Croft, R. Isserlin, J. B. Olsen, V. Fong, A. Emili, and A. W. Purcell, "Discovery of novel disease-specific and membrane-associated candidate markers in a mouse model of multiple sclerosis," *Molecular & Cellular Proteomics*, vol. 13, no. 3, pp. 679–700, 2014.

[53] S. Malhotra, N. Fissolo, M. Tintoré, A. C. Wing, J. Castilló, A. Vidal-Jordana, X. Montalban, and M. Comabella, "Role of high mobility group box protein 1 (hmgb1) in peripheral blood from patients with multiple sclerosis," *Journal of neuroinflammation*, vol. 12, no. 1, pp. 1–5, 2015.

[54] Y. N. Paudel, E. Angelopoulou, K. Bhuvan, C. Piperi, and I. Othman, "High mobility group box 1 (hmgb1) protein in multiple sclerosis (ms): Mechanisms and therapeutic potential," *Life sciences*, vol. 238, p. 116924, 2019.

[55] J. W. Peterson, L. Bö, S. Mörk, A. Chang, R. M. Ransohoff, and B. D. Trapp, "Vcam-1-positive microglia target oligodendrocytes at the border of multiple sclerosis lesions," *Journal of Neuropathology & Experimental Neurology*, vol. 61, no. 6, pp. 539–546, 2002.

[56] S. M. Seki, K. Posyniak, R. McCloud, D. A. Rosen, A. Fernández-Castañeda, R. M. Beiter, V. Serbulea, S. C. Nanziri, N. Hayes, C. Spivey, *et al.*, "Modulation of pkm activity affects the differentiation of th17 cells," *Science signaling*, vol. 13, no. 655, p. eaay9217, 2020.

[57] J. Gehrmann, R. B. Banati, M. L. Cuzner, G. W. Kreutzberg, and J. Newcombe, "Amyloid precursor protein (app) expression in multiple sclerosis lesions," *Glia*, vol. 15, no. 2, pp. 141–151, 1995.

[58] J. A. Matías-Guiu, C. Oreja-Guevara, M. N. Cabrera-Martín, T. Moreno-Ramos, J. L. Carreras, and J. Matías-Guiu, "Amyloid proteins and their role in multiple sclerosis. considerations in the use of amyloid-pet imaging," *Frontiers in neurology*, vol. 7, p. 53, 2016.

[59] L. Iparraguirre, M. Muñoz-Culla, I. Prada-Luengo, T. Castillo-Triviño, J. Olascoaga, and D. Otaegui, "Circular rna profiling reveals that circular rnas from anxa2 can be used as new biomarkers for multiple sclerosis," *Human Molecular Genetics*, vol. 26, no. 18, pp. 3564–3572, 2017.

[60] G. Scalabrino, "New epidermal-growth-factor-related insights into the pathogenesis of multiple sclerosis: Is it also epistemology?," *Frontiers in Neurology*, vol. 12, 2021.

[61] X. Lei, S. Cai, Y. Chen, J. Cui, Y. Wang, Z. Li, and Y. Li, "Down-regulation of interleukin 7 receptor (il-7r) contributes to central nervous system demyelination," *Oncotarget*, vol. 8, no. 17, p. 28395, 2017.

[62] L. F. Barcellos, S. Sawcer, P. P. Ramsay, S. E. Baranzini, G. Thomson, F. Briggs, B. C. Cree, A. B. Begovich, P. Villoslada, X. Montalban, *et al.*, "Heterogeneity at the hla-drb1 locus and risk for multiple sclerosis," *Human molecular genetics*, vol. 15, no. 18, pp. 2813–2824, 2006.

[63] A. S. Farias, F. Pradella, A. Schmitt, L. M. Santos, and D. Martins-de Souza, "Ten years of proteomics in multiple sclerosis," *Proteomics*, vol. 14, no. 4-5, pp. 467–480, 2014.

[64] H. Guan, P. S. Nagarkatti, and M. Nagarkatti, "Cd44 reciprocally regulates the differentiation of encephalitogenic th1/th17 and th2/regulatory t cells through epigenetic modulation involving dna methylation of cytokine gene promoters, thereby controlling the development of experimental autoimmune encephalomyelitis," *The Journal of Immunology*, vol. 186, no. 12, pp. 6955–6964, 2011.

[65] M. Gunnarsson, T. Stigbrand, and P. Jensen, "Decreased stability of a2-macroglobulin purified from patients with multiple sclerosis," *Acta neurologica scandinavica*, vol. 101, no. 4, pp. 270–272, 2000.

[66] X. Hu, G. Xiao, L. He, X. Niu, H. Li, T. Lou, Q. Hu, Y. Yang, Q. Xu, Z. Wei, *et al.*, "Sustained erbb activation causes demyelination and hypomyelination by driving necroptosis of mature oligodendrocytes and apoptosis of oligodendrocyte precursor cells," *Journal of Neuroscience*, vol. 41, no. 48, pp. 9872–9890, 2021.

[67] N. M. Moll, M. B. Cossoy, E. Fisher, S. M. Staugaitis, B. H. Tucky, A. M. Rietsch, A. Chang, R. J. Fox, B. D. Trapp, and R. M. Ransohoff, "Imaging correlates of leukocyte accumulation and cxcr4/cxcl12 in multiple sclerosis," *Archives of neurology*, vol. 66, no. 1, pp. 44–53, 2009.

[68] S. M. LeVine, "Albumin and multiple sclerosis," *BMC neurology*, vol. 16, no. 1, pp. 1–12, 2016.

[69] J. Barna, D. Dimén, G. Puska, D. Kovács, V. Csikós, S. Oláh, E. B. Udvari, G. Pál, and Á. Dobolyi, "Complement component 1q subcomponent binding

protein in the brain of the rat," *Scientific Reports*, vol. 9, no. 1, pp. 1–19, 2019.

[70] D. Moezzi, Y. Dong, R. W. Jain, B. M. Lozinski, S. Ghorbani, C. D'Mello, and V. Wee Yong, "Expression of antioxidant enzymes in lesions of multiple sclerosis and its models," *Scientific Reports*, vol. 12, no. 1, pp. 1–13, 2022.

[71] C. Cordiglieri, F. Baggi, P. Bernasconi, D. Kapetis, E. Faggiani, A. Consonni, F. Andreetta, R. Frangiamore, P. Confalonieri, C. Antozzi, *et al.*, "Identification of a gene expression signature in peripheral blood of multiple sclerosis patients treated with disease-modifying therapies," *Clinical Immunology*, vol. 173, pp. 133–146, 2016.

[72] E. Grajchen, E. Wouters, B. van de Haterd, M. Haidar, K. Hardonnière, T. Dierckx, J. Van Broeckhoven, C. Erens, S. Hendrix, S. Kerdine-Römer, *et al.*, "Cd36-mediated uptake of myelin debris by macrophages and microglia reduces neuroinflammation," *Journal of Neuroinflammation*, vol. 17, no. 1, pp. 1–14, 2020.

[73] D. C. Bullard, X. Hu, J. E. Adams, T. R. Schoeb, and S. R. Barnum, "p150/95 (cd11c/cd18) expression is required for the development of experimental autoimmune encephalomyelitis," *The American journal of pathology*, vol. 170, no. 6, pp. 2001–2008, 2007.

[74] M. Moransard, M. Bednar, K. Frei, M. Gassmann, and O. Ogunshola, "Erythropoietin reduces experimental autoimmune encephalomyelitis severity via neuroprotective mechanisms," *Journal of neuroinflammation*, vol. 14, no. 1, pp. 1–13, 2017.

[75] M. S. Natrajan, A. G. de la Fuente, A. H. Crawford, E. Linehan, V. Nunez, K. R. Johnson, T. Wu, D. C. Fitzgerald, M. Ricote, B. Bielekova, *et al.*, "Retinoid x receptor activation reverses age-related deficiencies in myelin debris phagocytosis and remyelination," *Brain*, vol. 138, no. 12, pp. 3581–3597, 2015.

[76] L. Szalardy, D. Zadori, E. Tanczos, M. Simu, K. Bencsik, L. Vecsei, and P. Klivenyi, "Elevated levels of ppar-gamma in the cerebrospinal fluid of

patients with multiple sclerosis," *Neuroscience Letters*, vol. 554, pp. 131–134, 2013.

[77] E. Fettah and A. DEMİR, "C-reactive protein/albumin ratio in patients with multiple sclerosis and its relationship with disease subtype and disability," *Journal of Surgery and Medicine*, vol. 4, no. 11, pp. 974–977, 2020.

[78] A. Datta, R. Gnanasambandan, S. Younes, T. Medha, R. Siva, H. Zayed, *et al.*, "Investigation of differentially expressed genes and dysregulated pathways involved in multiple sclerosis.," *Advances in protein chemistry and structural biology*, vol. 131, pp. 235–259, 2022.

[79] M. Wu, L. Xu, Y. Wang, N. Zhou, F. Zhen, Y. Zhang, X. Qu, H. Fan, S. Liu, Y. Chen, *et al.*, "S100a8/a9 induces microglia activation and promotes the apoptosis of oligodendrocyte precursor cells by activating the nf-κb signaling pathway," *Brain Research Bulletin*, vol. 143, pp. 234–245, 2018.

[80] C. T. Bever Jr and D. W. Garver, "Increased cathepsin b activity in multiple sclerosis brain," *Journal of the neurological sciences*, vol. 131, no. 1, pp. 71–73, 1995.

[81] A. Gaultier, X. Wu, N. Le Moan, S. Takimoto, G. Mukandala, K. Akassoglou, W. M. Campana, and S. L. Gonias, "Low-density lipoprotein receptor-related protein 1 is an essential receptor for myelin phagocytosis," *Journal of cell science*, vol. 122, no. 8, pp. 1155–1162, 2009.

[82] D. Pieragostino, P. Del Boccio, M. Di Ioia, L. Pieroni, V. Greco, G. De Luca, S. D'Aguanno, C. Rossi, D. Franciotta, D. Centonze, *et al.*, "Oxidative modifications of cerebral transthyretin are associated with multiple sclerosis," *Proteomics*, vol. 13, no. 6, pp. 1002–1009, 2013.

[83] M. G. Sotiropoulos and T. Chitnis, "Opposing and potentially antagonistic effects of bmp and tgf-β in multiple sclerosis: The "yin and yang" of neuroimmune signaling," *Journal of Neuroimmunology*, vol. 347, p. 577358, 2020.

[84] Y. M. Khaw, C. Cunningham, A. Tierney, M. Sivaguru, and M. Inoue, "Neutrophil-selective deletion of cxcr2 protects against cns neurodegeneration in a mouse model of multiple sclerosis," *Journal of neuroinflammation*, vol. 17, no. 1, pp. 1–12, 2020.

[85] T. Chitnis, "The role of cd4 t cells in the pathogenesis of multiple sclerosis," *International review of neurobiology*, vol. 79, pp. 43–72, 2007.

[86] J. Wang, J. Wang, J. Wang, B. Yang, Q. Weng, and Q. He, "Targeting microglia and macrophages: a potential treatment strategy for multiple sclerosis," *Frontiers in pharmacology*, vol. 10, p. 286, 2019.

[87] F. Ye, Y. Dai, T. Wang, J. Liang, X. Wu, K. Lan, and W. Sheng, "Transomics analyses revealed key epigenetic genes associated with overall survival in secondary progressive multiple sclerosis," *Journal of Neuroimmunology*, vol. 364, p. 577809, 2022.

[88] A. D. Rouillard, G. W. Gundersen, N. F. Fernandez, Z. Wang, C. D. Monteiro, M. G. McDermott, and A. Ma'ayan, "The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins," *Database*, vol. 2016, 2016.

[89] L. Thomas and L. A. Pasquini, "Galectin-3 prospects as a therapeutic agent for multiple sclerosis," *Neural Regeneration Research*, vol. 14, no. 8, p. 1380, 2019.

[90] G. Benedek, R. Meza-Romero, K. Jordan, Y. Zhang, H. Nguyen, G. Kent, J. Li, E. Siu, J. Frazer, M. Piecychna, *et al.*, "Mif and d-dt are potential disease severity modifiers in male ms subjects," *Proceedings of the National Academy of Sciences*, vol. 114, no. 40, pp. E8421–E8429, 2017.

[91] J. Melief, M. Orre, K. Bossers, C. G. van Eden, K. G. Schuurman, M. R. Mason, J. Verhaagen, J. Hamann, and I. Huitinga, "Transcriptome analysis of normal-appearing white matter reveals cortisol-and disease-associated gene expression profiles in multiple sclerosis," *Acta neuropathologica communications*, vol. 7, no. 1, pp. 1–19, 2019.

Appendix A

Data

The data underlying this thesis is available at Zenodo repository 7381894 (https://doi.org/10.5281/zenodo.7381894) in Supplementary Tables 1-7. The time we downloaded data is 2021-08-19 for IntAct, and 2021-11-02 for Bader-Lab Cell-Cell Interaction Database and CellTalkDB. In addition, the complete versions of tables A.23, A.24, A.25 and A.26 are available at the Zenodo repository in Supplementary Tables 8-11.

Table A.1: Demographic and clinical features of RRMS patients and healthy control (HC) subjects (EDSS: Expanded Disability Status Scale, SEM: Standart error of mean)

	RR-MS	HC
Subject number	8	7
Age (year)	52.75 ± 4.04 SEM	51.42 ± 3.76 SEM
Gender	6 Females	5 Females
	2 Males	2 Males
Disease duration (year)	17,62 ± 2,83 SEM	
EDSS	2.06 ± 0.27 SEM	
Immunomodulatory treatment, number		
Dimethyl fumarate	4	
Fingolimod	1	
Glatiramer acetate	1	
Interferon beta-1b	1	
Teriflunomide	1	

Table A.2: Primers used for qRT-PCR

HSP90AA1	
Forward:	CCCTGAATATCTGAACTTCATTAGAGG
Reverse:	GTCTTCGTGTATTCCAAGCTTTATG
HSP90B1	
Forward:	GAGGCTGAATCTTCTCCATTTG
Reverse:	CACAGCCTTTTCAATCTTGTCC
CALR	
Forward:	GAGGAGAAAGATAAAGGTTTGCAG
Reverse:	CACAGATGTCGGGACCAAAC
TFRC	
Forward:	CCATCAAGCTGCTGAATGAAA
Reverse:	GCGCTGTCTTTGACCTGAATC

Table A.3: Selected protein pairs by the ILP model for oligodendrocyte-macrophage network and for α=0.1 with PageRank based scores

Oligodendrocyte protein	**Macrophage protein**
ITGB1	HSP90AA1
HSP90AA1	TFRC

Table A.4: Selected protein pairs by the ILP model for oligodendrocyte-macrophage network and for α=0.2 with PageRank based scores

Oligodendrocyte protein	**Macrophage protein**
HMGB1	VCAM1
HSP90AA1	TFRC
ITGB1	ANXA2
FN1	PKM
APP	HSP90AA1

Table A.5: Selected protein pairs by the ILP model for oligodendrocyte-macrophage network and for α=0.4 with PageRank based scores

Oligodendrocyte protein	**Macrophage protein**
TFRC	P4HB
A2M	TFRC
HSP90B1	ALDOA
ANXA2	HSP90B1
HSP90AA1	CTSD
CTSD	PKM
PKM	HSP90AA1
APP	CALR
HMGB1	ITGAV
ITGB1	ANXA2
FN1	CD44

Table A.6: Selected protein pairs by the ILP model for oligodendrocyte-macrophage network and for α=0.6 with PageRank based scores

Oligodendrocyte protein	**Macrophage protein**
ERBB2	CTSD
HMGB1	ITGB3
APP	CD36
CXCR4	P4HB
HSP90AA1	PPARA
TFRC	HSP90B1
ALB	ALDOA
APOE	VCAM1
HSP90B1	ITGB2
A2M	ITGAV
ITGB1	HYOU1
CLU	CALR
CALR	ANXA2
ANXA2	LGALS3BP
CTSD	PKM
APOA1	KRT1
C1QBP	TFRC
KRT1	ANXA1
SOD1	HSP90AA1
FN1	ITGA3
PKM	CD44

Table A.7: Selected protein pairs by the ILP model for oligodendrocyte-macrophage network and for α=0.8 with PageRank based scores

Oligodendrocyte protein	**Macrophage protein**
NME1	TGFB1
EPOR	CXCR2
TRAC	CD4
RXRB	PPARA
IL7R	CALR
CXCR4	HMGB1
PPARG	CD36
ESRRA	TFRC
CRP	TNF
APP	ITGA6
ERBB2	ITGAV
SOD1	LGALS3
HLA-DRB1	PKM
CD44	ANXA1
ITGB1	ITGB3
PNP	VCAM1
S100A8	MIF
HMGB1	PPT1
APOE	ALDOA
HLA-A	P4HB
HSP90B1	ITGB2
TXLNA	HSP90AA1
ALB	S100A8
A2M	CD44
CLU	PRDX4
ANXA2	LGALS3BP
HSP90AA1	ANXA5
CTSD	HYOU1
PKM	CTSD
APOA1	KRT1
KRT1	ALB
CTSB	ANXA2
CALR	APOB
LRP1	HSP90B1
TTR	APOA1
FN1	ITGA3

Table A.8: Selected protein pairs by the ILP model for oligodendrocyte-T-cell network and for α=0.1 with PageRank based scores

Oligodendrocyte protein	**T-cell protein**
APP	APP
HSP90AA1	EGFR
FN1	ITGB1

Table A.9: Selected protein pairs by the ILP model for oligodendrocyte-T-cell network and for α=0.2 with PageRank based scores

Oligodendrocyte protein	**T-cell protein**
CLU	EGFR
PKM	ANXA2
ANXA2	PKM
ITGB1	FN1
HSP90AA1	HSP90AA1
APP	APP
FN1	ITGB1

Table A.10: Selected protein pairs by the ILP model for oligodendrocyte-T-cell network and for α=0.4 with PageRank based scores

Oligodendrocyte protein	**T-cell protein**
CLU	APP
APP	TFRC
APOA1	VCP
HSP90B1	ITGB1
CD44	HSP90AA1
HSP90AA1	ITGA4
ALB	ALDOA
APOE	ANXA2
KRT1	EGFR
A2M	ITGA5
BSG	CLU
CALR	HSP90B1
TFRC	BSG
ITGB1	C1QBP
ANXA2	PKM
CTSD	FN1
FN1	ITGAV
PKM	CD44

Table A.11: Selected protein pairs by the ILP model for oligodendrocyte-T-cell network and for α=0.6 with PageRank based scores

Oligodendrocyte protein	**T-cell protein**
SMC3	ITGB2
LAMB1	FN1
GPC1	APP
HSP90AA1	ITGAV
ESRRA	ALDOA
ERBB2	ITGA5
COL1A2	P4HB
APP	TNF
ITGA6	LGALS3BP
APOA1	VCP
ANXA1	BSG
ANXA2	PNP
PNP	PKM
ALB	TFRC
GPC4	HSP90AA1
VCP	SLC1A5
CLU	IGF2R
CD44	HYOU1
A2M	ATP1B3

Table A.12: Selected protein pairs by the ILP model for oligodendrocyte-T-cell network and for α=0.8 with PageRank based scores

Oligodendrocyte protein	T-cell protein
BCAN	PNP
LAMB1	NR3C1
LUM	A2M
PTPRC	CD44
LRP1	PSAP
ITGB8	FN1
CD44	MIF
TRAC	EGFR
RXRA	TNF
LGALS3	SOD1
ITGB1	TGFB1
IL7R	RPSA
HMGB1	ITGB2
GPC1	NME1
ESRRA	ALDOA
EPOR	TRAF2
COL2A1	P4HB
APP	CD36
ERBB2	ITGAV
ITGA6	LGALS3BP
APOA1	VCP
HLA-DRB1	PKM
ANXA1	BSG
ITGB4	TFRC
PNP	HMGB1
ALB	SERPINA1
TF	S100A8
KRT1	CTSD
PKM	PPT1
VCP	HSD17B12
GPC4	C1QBP

Oligodendrocyte protein	T-cell protein
CALU	HYOU1
HSP90B1	HSPA13
CALR	CALU
HLA-A	CALR
TXLNA	HSP90AA1
SMC3	RUNX1
GSN	HNRNPM
TFRC	GUSB
CLU	ITGA3
ATP1B3	P3H1
LGALS3BP	FBN1
SERPINA1	IGF2R
A2M	LAMA5
APOB	ANXA1
APOE	SLC1A5
APOC3	ALB
ALDOA	TOR1A
CTSD	ANXA2
C1QBP	M6PR
BSG	ATP1B3
HSP90AA1	PCYOX1
P4HB	SDF2L1
C3	KRT1
LGALS1	AP2M1
ANXA2	ITGA5
TNC	CLU
TTR	SDF4
HLA-DRA	HSP90B1
BGN	APP
FN1	ITGA4
COL1A2	ITGB1

Table A.13: Selected protein pairs by the ILP model for oligodendrocyte-macrophage network and for α=0.1 with Betweenness based scores

Oligodendrocyte protein	**Macrophage protein**
HSP90AA1	HSP90AA1
ITGB1	TFRC

Table A.14: Selected protein pairs by the ILP model for oligodendrocyte-macrophage network and for α=0.2 with Betweenness based scores

Oligodendrocyte protein	**Macrophage protein**
APP	HSP90AA1
FN1	ANXA2
ITGB1	PKM
HMGB1	TFRC
HSP90AA1	VCAM1

Table A.15: Selected protein pairs by the ILP model for oligodendrocyte-macrophage network and for α=0.4 with Betweenness based scores

Oligodendrocyte protein	**Macrophage protein**
FN1	ITGAV
APP	CTSD
PKM	TFRC
CTSD	ANXA2
CLU	P4HB
A2M	HYOU1
HMGB1	HSP90AA1
HSP90AA1	CD44

Table A.16: Selected protein pairs by the ILP model for oligodendrocyte-macrophage network and for α=0.6 with Betweenness based scores

Oligodendrocyte protein	**Macrophage protein**
HSP90AA1	ITGB3
FN1	ITGA3
APP	CTSD
KRT1	ANXA1
C1QBP	ITGAV
ITGB1	PKM
PKM	HSP90B1
CTSD	ANXA2
ANXA2	HSP90AA1
HSP90B1	CALR
A2M	LGALS3BP
CLU	TFRC
TFRC	HYOU1
APOE	VCAM1
S100A8	KRT1
ALB	ALDOA
ALDOA	CD44
CXCR4	P4HB
HMGB1	ITGB2
PPARG	CD36
ERBB2	PPARA

Table A.17: Selected protein pairs by the ILP model for oligodendrocyte-macrophage network and for α=0.8 with Betweenness based scores

Oligodendrocyte protein	**Macrophage protein**
HSP90AA1	ITGB3
S100A8	ITGB2
FN1	ITGA3
LRP1	HSP90B1
CTSB	ANXA2
KRT1	HMGB1
C1QBP	ITGAV
ITGB1	TFRC
C3	KRT1
PKM	CTSD
CALR	PRDX4
CLU	P4HB
HSP90B1	ANXA5
TTR	ALB
ALB	APOA1
A2M	LGALS3BP
CD44	HYOU1
ERBB2	CD44
TXLNA	HSP90AA1
ANXA2	CD4
APOE	VCAM1
CTSD	S100A8
HMGB1	PPT1
PNP	ANXA1
HLA-DRB1	PKM
APOA1	APOB
RXRA	PPARA
SOD1	LGALS3
APP	ITGA6
CRP	TNF
ESRRA	ALDOA
GRM3	TGFB1
PPARG	CD36
CXCR4	MIF
IL7R	CALR
EPOR	CXCR2

Table A.18: Selected protein pairs by the ILP model for oligodendrocyte-T-cell network and for α=0.1 with Betweenness based scores

Oligodendrocyte protein	**T-cell protein**
APP	HSP90AA1
ITGB1	EGFR
FN1	APP

Table A.19: Selected protein pairs by the ILP model for oligodendrocyte-T-cell network and for α=0.2 with Betweenness based scores

Oligodendrocyte protein	**T-cell protein**
FN1	ITGB1
ITGB1	FN1
APP	HSP90AA1
HSP90AA1	PKM
CLU	ANXA2
PKM	EGFR
ANXA2	APP

Table A.20: Selected protein pairs by the ILP model for oligodendrocyte-T-cell network and for α=0.4 with Betweenness based scores

Oligodendrocyte protein	**T-cell protein**
FN1	ITGAV
KRT1	FN1
TFRC	C1QBP
PKM	ITGA5
CTSD	CLU
CLU	ANXA2
A2M	BSG
ANXA2	EGFR
APOE	PKM
ALB	ALDOA
HSP90AA1	ITGA4
CD44	HSP90AA1
BSG	HSP90B1
CALR	ITGB1
HSP90B1	CD44
APOA1	APP
APP	TFRC
ITGB1	VCP

Table A.21: Selected protein pairs by the ILP model for oligodendrocyte-T-cell network and for α=0.6 with Betweenness based scores

Oligodendrocyte protein	**T-cell protein**
FN1	ITGA4
APP	CTSD
KRT1	TFRC
ANXA2	ITGA5
ANXA1	KRT1
ITGB1	AP2M1
PKM	ITGAV
APOA1	CLU
C3	FN1
TFRC	BSG
C1QBP	ATP1B3
P4HB	LGALS3BP
ALB	SLC1A5
ITGB4	ANXA1
BSG	P3H1
A2M	ITGA3
ATP1B3	HYOU1
CTSD	IGF2R
GSN	ANXA2
ERBB2	CD44
HSP90B1	ITGB2
GPC4	C1QBP
SMC3	VCP
APOE	PKM
CLU	ALB
HSP90AA1	PNP
CD44	HSP90AA1
CALR	ITGB1
TTR	A2M
VCP	CALR
COL1A2	P4HB
ESRRA	ALDOA
PNP	HSP90B1
GPC1	APP
ITGA6	EGFR

Table A.22: Selected protein pairs by the ILP model for oligodendrocyte-T-cell network and for α=0.8 with Betweenness based scores

Oligodendrocyte protein	T-cell protein
LAMB1	ITGB1
APP	TGFB1
BGN	APP
CALR	NR3C1
TTR	SDF4
TNC	CLU
COL2A1	FN1
ANXA2	RUNX1
ALDOA	ITGA4
ANXA1	TFRC
HMGB1	ITGA5
LGALS1	KRT1
P4HB	SDF2L1
HSP90AA1	PCYOX1
TFRC	BSG
C1QBP	M6PR
C3	HNRNPM
PKM	TOR1A
APOC3	ALB
APOA1	SLC1A5
APOB	AP2M1
ITGB4	ANXA1
BSG	GUSB
ITGB1	CTSD
A2M	CD36
SERPINA1	HYOU1
CD44	P3H1
LGALS3BP	LAMA5
ATP1B3	IGF2R
CLU	ITGA3
CALU	HSD17B12

Oligodendrocyte protein	T-cell protein
TXLNA	HSP90AA1
TF	SERPINA1
HSP90B1	HSPA13
HLA-A	CALU
GPC4	C1QBP
APOE	ANXA2
CTSD	HMGB1
KRT1	S100A8
GSN	ITGAV
LRP1	CALR
HLA-DRB1	VCP
HLA-DRA	PKM
ITGA6	LGALS3BP
ERBB2	PSAP
ALB	TNF
COL1A2	P4HB
EPOR	TRAF2
ESRRA	ALDOA
GPC1	NME1
IL7R	RPSA
LGALS3	SOD1
PNP	HSP90B1
RXRA	ATP1B3
TRAC	EGFR
FN1	FBN1
PTPRC	CD44
VCP	PPT1
LUM	A2M
SMC3	ITGB2
ITGB8	MIF
BCAN	PNP

Table A.23: Top 40 contact protein pairs based on individual overall PageRank based scores in the oligodendrocyte-macrophage network

Oligodendrocyte protein	**Macrophage protein**	**Overall score**
APP	PKM	1.8093
APP	HSP90AA1	1.8084
HSP90AA1	PKM	1.8059
HSP90AA1	HSP90AA1	1.8051
FN1	PKM	1.7986
FN1	HSP90AA1	1.7978
APP	TFRC	1.7974
HSP90AA1	TFRC	1.7940
FN1	TFRC	1.7867
APP	ANXA2	1.7499
HSP90AA1	ANXA2	1.7466
FN1	ANXA2	1.7393
ITGB1	PKM	1.7367
ITGB1	HSP90AA1	1.7359
ITGB1	TFRC	1.7248
ITGB1	ANXA2	1.6774
HSP90AA1	ALDOA	1.6769
FN1	ALDOA	1.6696
HSP90AA1	CD44	1.6553
PKM	PKM	1.6525
PKM	HSP90AA1	1.6517
HSP90AA1	ITGAV	1.6488
FN1	CD44	1.6481
HMGB1	PKM	1.6480
HMGB1	HSP90AA1	1.6472
FN1	ITGAV	1.6415
PKM	TFRC	1.6406
APP	CTSD	1.6365
HMGB1	TFRC	1.6361
HSP90AA1	CTSD	1.6331
APP	P4HB	1.6286
HSP90AA1	VCAM1	1.6269
FN1	CTSD	1.6258
HSP90AA1	P4HB	1.6252
HSP90AA1	HSP90B1	1.6227
ANXA2	PKM	1.6212
ANXA2	HSP90AA1	1.6204
ANXA2	TFRC	1.6093
ITGB1	ALDOA	1.6077
TFRC	PKM	1.6047

Table A.24: Top 40 contact protein pairs based on individual overall PageRank based scores in the oligodendrocyte-T-cell network

Oligodendrocyte protein	**T-cell protein**	**Overall score**
FN1	EGFR	1.9948
FN1	APP	1.9494
FN1	HSP90AA1	1.9233
FN1	FN1	1.9138
FN1	ITGB1	1.9005
FN1	PKM	1.8716
FN1	TFRC	1.8559
FN1	ANXA2	1.8365
APP	EGFR	1.8252
FN1	ALDOA	1.8123
FN1	CLU	1.8113
FN1	C1QBP	1.8112
ITGB1	EGFR	1.7995
FN1	VCP	1.7893
APP	APP	1.7798
FN1	CTSD	1.7719
FN1	CD44	1.7694
FN1	LGALS3BP	1.7571
APP	HSP90AA1	1.7538
FN1	ANXA1	1.7527
FN1	AP2M1	1.7468
APP	FN1	1.7443
FN1	HMGB1	1.7423
FN1	HNRNPM	1.7411
FN1	KRT1	1.7395
FN1	ITGAV	1.7379
HSP90AA1	EGFR	1.7371
FN1	SLC1A5	1.7346
FN1	ITGA4	1.7332
FN1	ALB	1.7291
ITGB1	HSP90AA1	1.7281
FN1	RPSA	1.7246
FN1	IGF2R	1.7241
FN1	SMC3	1.7236
ITGB1	FN1	1.7185
FN1	ITGA5	1.7104
FN1	P3H1	1.7065
ITGB1	ITGB1	1.7053
APP	PKM	1.7021
FN1	SOD1	1.7002

Table A.25: Top 40 contact protein pairs based on individual overall Betweenness based scores in the oligodendrocyte-macrophage network

Oligodendrocyte protein	**Macrophage protein**	**Overall Score**
APP	PKM	1.8714
APP	HSP90AA1	1.8346
APP	TFRC	1.8318
HSP90AA1	PKM	1.8192
APP	ANXA2	1.8022
HSP90AA1	HSP90AA1	1.7824
HSP90AA1	TFRC	1.7797
FN1	PKM	1.7773
ITGB1	PKM	1.7731
HSP90AA1	ANXA2	1.7500
FN1	HSP90AA1	1.7405
FN1	TFRC	1.7378
ITGB1	HSP90AA1	1.7363
ITGB1	TFRC	1.7336
FN1	ANXA2	1.7081
ITGB1	ANXA2	1.7039
HMGB1	PKM	1.6925
PKM	PKM	1.6877
HSP90AA1	ALDOA	1.6694
APP	CTSD	1.6603
HMGB1	HSP90AA1	1.6558
HMGB1	TFRC	1.6530
PKM	HSP90AA1	1.6509
PKM	TFRC	1.6481
APP	P4HB	1.6441
ALDOA	PKM	1.6424
ANXA2	PKM	1.6323
HSP90AA1	CD44	1.6290
FN1	ALDOA	1.6275
ERBB2	PKM	1.6269
HSP90B1	PKM	1.6259
HMGB1	ANXA2	1.6233
ITGB1	ALDOA	1.6233
HSP90AA1	ITGAV	1.6219
PKM	ANXA2	1.6185
CLU	PKM	1.6142
HSP90AA1	CTSD	1.6082
APP	CALR	1.6078
ALDOA	HSP90AA1	1.6057
TFRC	PKM	1.6044

Table A.26: Top 40 contact protein pairs based on individual overall Betweenness based scores in the oligodendrocyte-T-cell network

Oligodendrocyte protein	**T-cell protein**	**Overall score**
FN1	EGFR	1.9650
FN1	APP	1.9267
FN1	HSP90AA1	1.9166
FN1	ITGB1	1.8813
FN1	FN1	1.8794
FN1	PKM	1.8755
APP	EGFR	1.8458
FN1	TFRC	1.8279
FN1	ANXA2	1.8208
ITGB1	EGFR	1.8158
FN1	ALDOA	1.8099
APP	APP	1.8075
FN1	C1QBP	1.7988
FN1	CLU	1.7988
APP	HSP90AA1	1.7974
FN1	VCP	1.7767
ITGB1	HSP90AA1	1.7674
APP	FN1	1.7603
APP	PKM	1.7564
HSP90AA1	EGFR	1.7487
FN1	CTSD	1.7463
FN1	CD44	1.7391
FN1	AP2M1	1.7348
FN1	LGALS3BP	1.7345
ITGB1	ITGB1	1.7321
ITGB1	FN1	1.7303
FN1	ANXA1	1.7266
ITGB1	PKM	1.7264
FN1	HNRNPM	1.7208
FN1	KRT1	1.7201
FN1	SLC1A5	1.7186
FN1	HMGB1	1.7112
FN1	RPSA	1.7104
HSP90AA1	APP	1.7104
FN1	ALB	1.7092
APP	TFRC	1.7088
FN1	ITGAV	1.7083
FN1	IGF2R	1.7050
APP	ANXA2	1.7016
FN1	SMC3	1.7011

Table A.27: The most frequent proteins in the top 20 intracellular contributors for the selected highest scoring 10 proteins

T-cell protein	**Oligodendrocyte protein**	**Macrophage protein**
YWHAZ	YWHAZ	YWHAZ
GRB2	CDC5L	GRB2
CUL3	HTT	HSPB1
COPS5	GRB2	HSP90AB1
DISC1	DLD	HSPA8
HSCB	COPS5	HSPA5
LRRK2	SNW1	MYH9
	PRNP	ACTB
	CAND1	COPS5
	OTUB1	H2AX

Appendix B

Code

BriFin is available at https://github.com/BilkentCompGen/brifin.