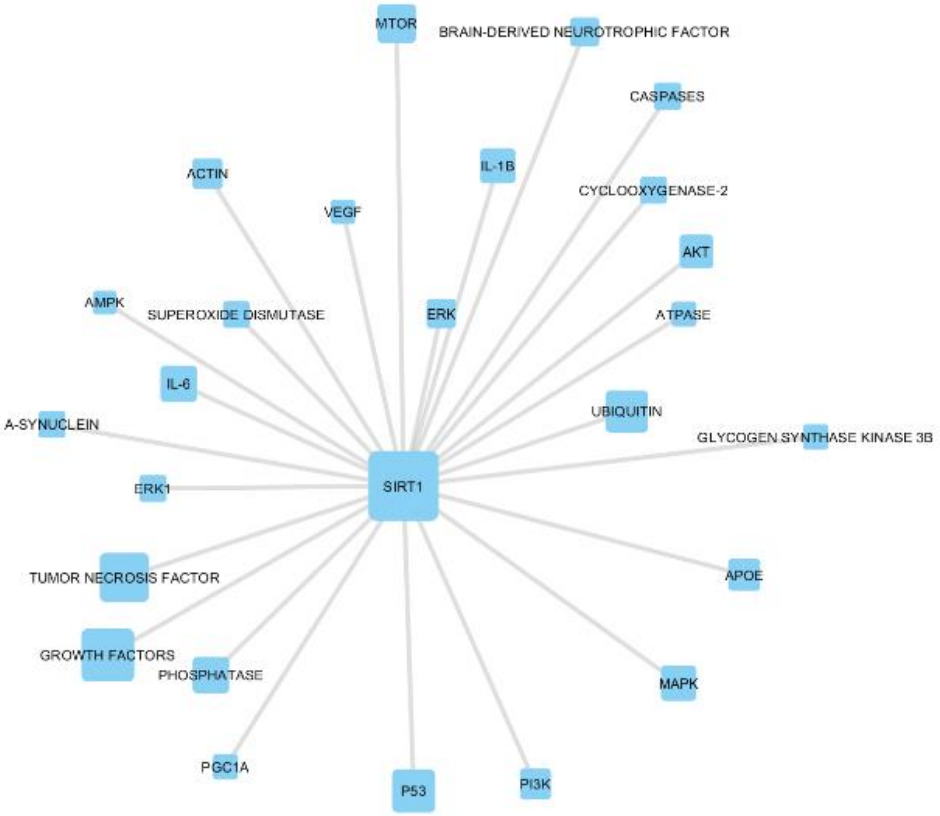
Universiteit Leiden
The Netherlands

Leiden Institute of Chemistry

Investigating Sirtuins in Alzheimer's Disease Using Novel Computational Techniques



Colloquium Essay

March 2020

Student: B.J. Molenaar (s1121146)

Supervisor: Dr M. Palmblad

Mentor: Dr A. Pandit

Jury Members: Dr R. G. Boots and Dr M.E. Artola Perez de Azanza

Word Count: 4684

6 EC

Life Science and Technology

Abstract

The most prevalent neurodegenerative disease is Alzheimer's Disease (AD). No breakthrough treatment has been found yet, indicating the importance of researching this disease further. However, an enormous amount of research is performed each year (15,018 articles in 2018 alone), which makes it impossible for humans to analyse and group all the gathered information. Therefore, tools should be developed to summarize this vast amount of information. In this work, we used a novel and reusable text-mining based approach to investigate recent protein interactions with SIRT1 relating to AD. A protein network was created consisting of co-occurrences and three novel proteins interacting with SIRT1 were found. SIRT1 is upregulated by FOXQ1, which prevents apoptosis. SIRT1 is also upregulated by S1PR1, however, there are also contradictory findings. Lastly, CLDN5, which is important in the integrity of the blood-brain barrier (BBB), is regulated by SIRT1. These discoveries show that using this text-mining approach we were able to find recent interactions of SIRT1 in AD.

Introduction

Ageing is the foremost risk factor for neurodegenerative diseases. The most prevalent neurodegenerative disease is Alzheimer's Disease (AD). In 2006, the European prevalence of AD was 7.21 million.[1] However, due to the rapid ageing of the population, it is forecasted that by 2050, the prevalence of AD will be more than doubled.[1] To prevent this growth, it is important to understand the ageing process that causes AD.

However, understanding AD and the role of ageing in AD's development is shown to be difficult albeit the immense amount of research that has been done. Only in 2018, 15,018 articles regarding AD have been published and recorded in the Europe PMC database. This number means that you have to read 41 papers a day if you want to be up to date on all the AD research. Therefore, it is important that tools are being developed to analyse and summarize this fast amount of papers.

Currently, developments in text-mining make it possible to extract meaningful information from the enormous amount of scientific literature. Even without human supervision, Tshitoyan e.a. (2019)[2] showed the possibility to recommend materials for functional applications using 3.3 million scientific abstracts. Additionally, Palmblad (2019)[3] used text-mining to predict the physiochemical properties of chemical entities. Lastly, STRING database, which is a database that aims to collect all publicly available protein-protein interaction information, uses text-mining as one of its seven information sources.[4] However, despite its ease of use, such a database is limited to the implemented functions. These examples highlight the opportunities in extracting meaningful information using text-mining and the potential of using these techniques in understanding diseases such as AD.

In this project, we use a text-mining process based on Guler's e.a. (2016)[5] work to create a protein network. More specifically, we create a protein co-occurrence network of the Sirtuin protein family with the relation to AD. Based on this method, we found that Silent mating type information regulation 2 homolog 1 (SIRT1), an NAD^+-dependent deacetylase, has interesting connections in the co-occurrence network. This work describes the process of creating the protein co-occurrence network, the analyses of the created network and uses the network to find novel interactions between SIRT1 and other proteins.

Material and Methods

Literature Search

The method used for this colloquium is based on the work of Guler e.a. (2016)[5] and the STRING database.[4] However, in this work, the method is based on two web-based services. This method starts with a literature search using the query: (sirtuin OR Sirtuins OR SIRT1 OR SIRT2 OR SIRT3 OR SIRT4 OR SIRT5 OR SIRT6 OR SIRT7) AND (Alzheimer OR alzheimer), which is used for one API call to the Europe PMC searchPublications RESTful service. This call results in a list of PubMed article IDs with their corresponding metadata. Using this list, the annotated proteins were retrieved using the getAnnotations service. This service uses text-mining to annotate information in articles, including named proteins. After these searches, a co-occurrence edge-list was made (Figure 1). For each co-occurrence, the average publication year and the weight (Equation 1) was calculated. These two values were used to filter the co-occurrence edge list. Additionally, the PubMed article IDs, which are retrievable from the meta-data and used later for writing the review, were also added. Then, the co-occurrences that only were found once were discarded. Lastly, the co-occurrence edge list was stored in a CSV file using the '\t' or 'tab' delimiter. This whole procedure is made in a Jupiter Notebook (https://github.com/Mees-Molenaar/colloquium).

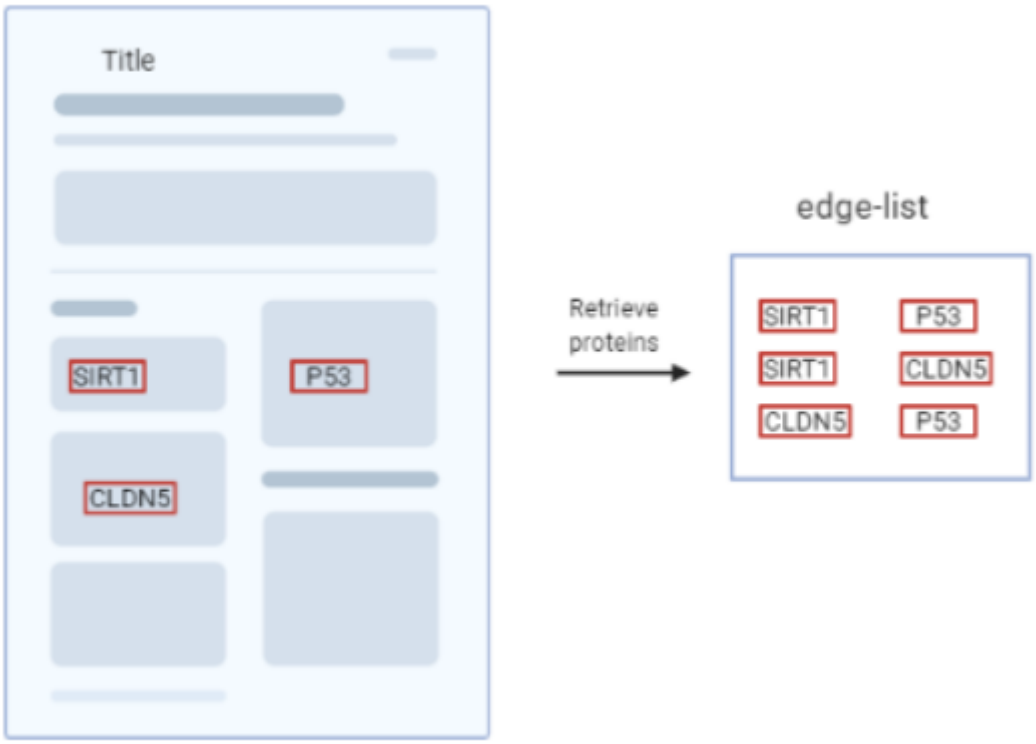


Figure 1: Schematic overview of retrieving the proteins from an article. After the retrieval, the proteins are stored in a co-occurrence edge-list. In this edge-list, the average publication year, a weight factor and the PubMed article IDs are added.

Protein Network

After creating the co-occurrence edge list, the data was visualized using Cytoscape. The CSV file was imported as a network. Be aware of setting the delimiter to '\t' or 'tab' and not including ',', since some protein names contain a comma. Gene Symbol 1 was used as the source node and Uniprot ID 1 as the source attribute. Gene symbol 2 was used as the target node and Uniprot ID 2 as the target attribute. The average year, weight and article IDs were used as edge attributes. Then the network was trimmed using edit→remove self-loops and edit→remove duplicate edges. After trimming, the network analyser was used by using Tools→NetworkAnalyzer→Network Analysis→Analyze Network while treating the network as undirected. For the style, the size was set to continuous mapping based on degree starting from 10 and ending on 30. The edge opacity was set to 66 (Figure 2).

After setting up the protein network, the network was filtered accordingly. For the most degrees protein network, the network was filtered on the degree including every node with a degree higher than 2,500 (Figure 3). For the finale protein network with the four interesting proteins, the network was filtered for multiple edge attributes: the shared name contains SIRT1, the average year is 2018 and the weight is bigger than 0.1 (Figure 4).

Literature Review

For writing the review, the highest degree network (Figure 3) was used to give a general overview of AD. Additionally, the recent co-occurrence network (Figure 4) was used to go in-depth on the relation between *SIRT1* and the four found proteins. For each edge (co-occurrence), the PubMed article IDs were also saved. These articles were used to get an understanding of the possible biological relationship between the co-occurrence. When this biological relationship was found to exist, the search was broadened and extended using the Europe PMC database.

Results

Literature Search

The literature search in the Europe PMC database resulted in 1,353 articles*. In these articles, a total of 225,800 annotated proteins were found. However, these proteins were sometimes mentioned multiple times in the same article. Therefore, these duplicates were removed. Additionally, some articles were not annotated correctly, which were also removed and this clean-up resulted in 1,309 correctly annotated articles. Furthermore, articles that had over 1,000 annotated proteins have been discarded. This threshold left 1,305 articles, which resulted in 4,174,306 co-occurrences in an edge list. For each co-occurrence, the average publication date and weight were calculated. This weight was calculated using equation 1, where N is the total amount of proteins per article. Then, the co-occurrences that were mentioned only once were removed. This deletion resulted in 634,713 co-occurrences in the edge list. With this edge list, the protein network is build using Cytoscape.

$$\sum \frac{2}{(N(N-1))} \tag{1}$$

Protein Network

The edge list of the full network has 6,595 nodes and 611,400 edges (Figure 1). The network is in a typical hairball structure, which is not interpretable. Therefore, firstly, only the nodes with the highest degree (>2,500) are shown (Figure 2). This network has a total of 26 nodes and 377 edges. Furthermore, these nodes contain the proteins that are most connected with AD such as the mammalian target of rapamycin (*mTOR)*, tumour protein p53 (*P53)*, apolipoprotein E (*APOE)* and, of course, *SIRT1* (Table 1). Interestingly, one big player, amyloid precursor protein, (*APP)*, is missing from the entire work. *APP* is probably not annotated by Europe PMC since it is also the abbreviation of application, app. This abbreviation results in a lot of hits that are not meaningful, therefore, Europe PMC has chosen to discard in from the text-mining results. Furthermore, *amyloid-β* (Aβ) is also not shown in the network, which might be due to Aβ being a peptide. Lastly, also microtubule-associated tau protein (*MAPT*) is missing in the network with the highest degree, however, it is found in the network with a degree of 1,931. These results show that cautiousness is necessary when analysing protein networks via this procedure.

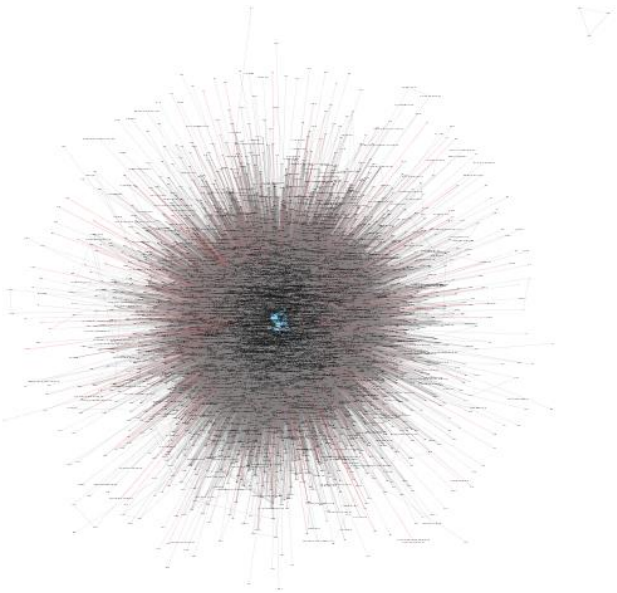


Figure 2: The full protein interaction network for Sirtuins in AD. The network is created from the co-occurrence edge-list and is visualized using Cytoscape. The nodes are represented by blue squares containing the gene name and an is edge is a co-occurrence. The size of the nodes is mapped to the degree.

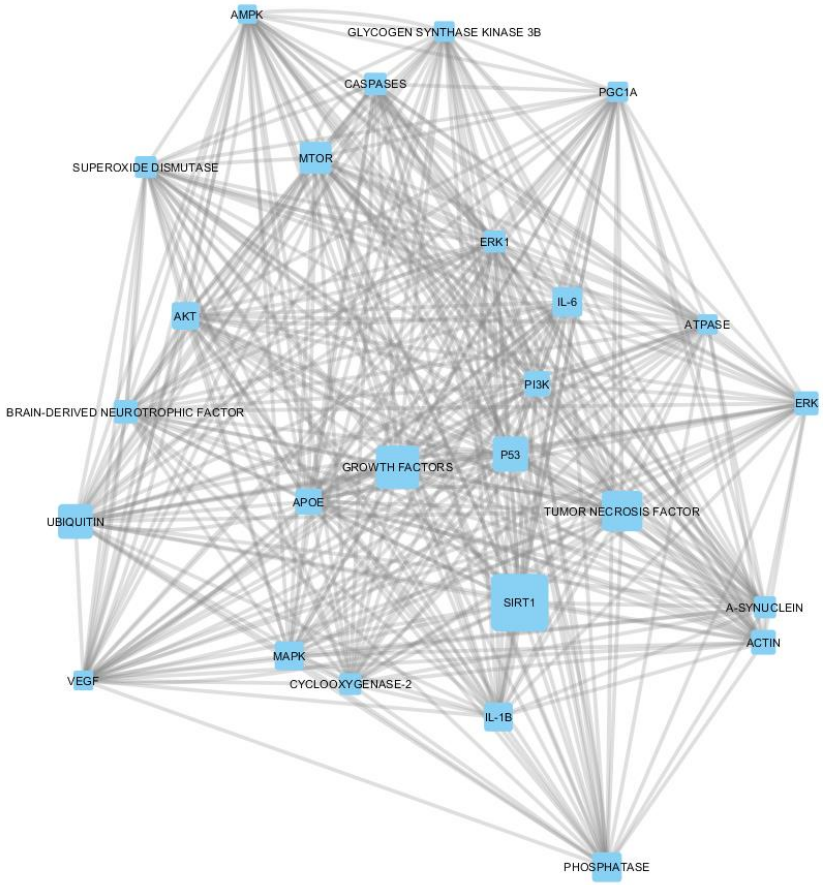


Figure 3: The protein interaction network of Sirtuins in AD trimmed down to show only the nodes with the highest degree (>2500). The network is visualized using Cytoscape where the nodes are represented by blue squares containing the gene name and an edge is a co-occurrence. The size of the nodes is mapped to the degree.

Table 1: Table containing the nodes with the highest degree (>2500) of the protein interaction network of Sirtuins in AD (Figure 3). The table is sorted based on degree and the gene name, the degree and the Uniprot ID is given.

Gene Name	Degree	Uniprot ID
SIRT1	5442	Q96EB6
GROWTH FACTORS	4360	Q8V307
TUMOR NECROSIS FACTOR	4124	P59684
P53	3742	Q29537
UBIQUITIN	3703	P69310
MTOR	3486	P42345
IL-6	3336	P26892
PHOSPHATASE	3325	Q9NJE9
MAPK	3264	O42781
IL-1B	3249	P09428
AKT	3154	Q8INB9
APOE	3056	P0DKW5
PI3K	2995	P54673
ACTIN	2947	Q05214
ERK	2888	P29323
BRAIN-DERIVED NEUROTROPHIC FACTOR	2859	Q1X706
CASPASES	2763	O01382
ALPHA-SYNUCLEIN	2740	Q3I5G7
ERK1	2738	P27361
CYCLOOXYGENASE-2	2735	O62698
SUPEROXIDE DISMUTASE	2718	E3YBA4
GLYCOGEN SYNTHASE KINASE 3B	2632	P49841
PGC1A	2622	Q865B7
ATPASE	2613	A3DIJ8
VEGF	2568	P15691
AMPK	2521	Q9Y478

Since the focus on this project lies on the effect of SIRT1 on AD, a protein network was made consisting of co-occurrences which mentioned SIRT1. This protein network consisted of 4653 nodes and 5442 edges. Furthermore, we were interested in recent findings. Thus, we further trimmed the network by focusing on interactions of which the average publication date is 2018. We have chosen 2018 since it might be that the Europe PMC database did not annotate all of the articles in 2019 and 2020 yet. This protein network consisted of 677 nodes and 692 edges. This network is still very crowded, therefore, we trimmed it down further using the calculated weight (Equation 1). The threshold used was 0.1, which resulted in 4 co-occurrences with *SIRT1*: Protein Kinase AMP-Activated Catalytic Subunit Alpha 2 (*PRKAA2)*, Forkhead Box Q1 (*FOXQ1)*, Sphingosine-1-phosphate (*S1P)* RECEPTOR-1 *(S1PR1)* and Claudin-5 *(CLDN5)* (Figure 4; Table 2).

However, *PRKAA2* has a synonym *AMPK*. This synonym is also seen in the big protein network (Table 1). Furthermore, the average publication year of that protein is 2015, which is not recent enough to focus on for this project. Therefore, this protein is not discussed in-depth. This result shows that naming proteins in a consistent way is very important for research based on text-mining. Or, the software should take into account different naming and add the results together.

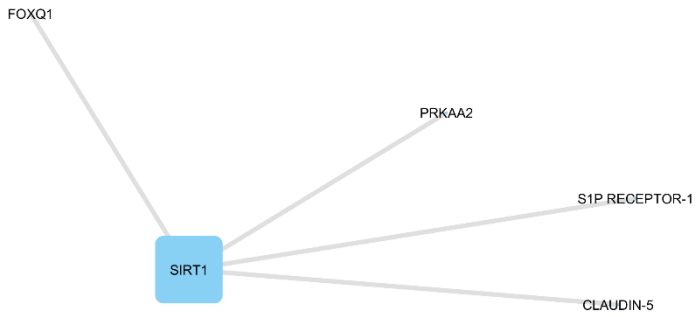


Figure 4: The protein interaction network of Sirtuins in AD. This network is limited to the SIRT1 gene, edges that have the average year of 2018 and a weight that is higher than 0.1. The network is visualized using Cytoscape where the nodes are represented by blue squares containing the gene name and an edge is a co-occurrence. The size of the nodes is mapped to the degree.

Table 2: Table containing the nodes connected to the edges with the highest weight, the average year of 2018 and connected to SIRT1 (Figure 4). The weight is calculated by equation 1 and the gene name, the degree and Uniprot ID are given. The table is sorted based on weight.

Gene Name	Degree	Uniprot ID	weight
CLAUDIN-5	341	Q2HJ22	1.00
PRKAA2	88	P54646	0.17
S1P RECEPTOR-1	331	Q5E9P3	0.17
FOXQ1	2	Q9C009	0.10
SIRT1	5442	Q96EB6	-

Literature Review

In this review, the influence of SIRT1 on AD will be discussed. However, first, the key proteins in regards to AD will be introduced (Figure 5). This introduction is lightly based on the highest degree protein network and is necessary to understand the cause of AD. An important key player in AD, *APP*, was missing from the made network, however, due to its importance, it is included in this review. Then, SIRT1 will be introduced and described. Lastly, the three interesting proteins *FOXQ1*, *S1PR1* and *CLDN5* will be discussed in-depth.

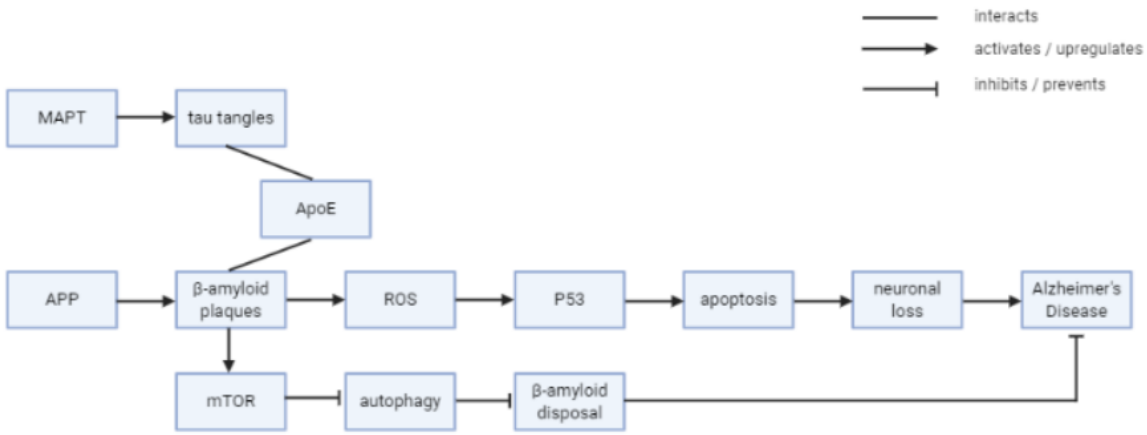


Figure 5: Schematic summary of the key players important in the development of Alzheimer's Disease.

APP: A Missing Key Player

One of the main factors in AD is the formation of β-amyloid (Aβ) plaques. The importance of Aβ plaques was first found in familial AD patients, who had a mutation in the Aβ precursor protein (APP).[6] Mutated APPs are wrongly processed and accumulate in a cell, which causes an increase in reactive oxygen species (ROS).[7] This ROS damages brain cells, which results in apoptosis. Thus, erroneous processed APP results in Aβ plaques, which cause apoptosis and neuronal loss in AD.

The Found Key Players

P53

One of the pathways that result in apoptosis is regulated by tumour protein p53. P53 is the most researched protein[8] and promotes apoptosis in response to cellular stress.[9] For example, Aβ plaques increase ROS production, which causes DNA damage.[10] As a response to this damage, p53 is activated by phosphorylation. This phosphorylation starts a chain reaction leading to apoptosis. Additionally, in AD brains, the amount of p53 was increased compared to control brains.[11] This rise indicates the role of p53 in AD's neuronal loss.

mTOR

Another role that p53 and Aβ plaques have in AD is connected to the mammalian target of rapamycin (mTOR). mTOR is an important regulator of cell growth and proliferation.[12] Furthermore, an important process that is negatively regulated by mTOR is autophagy,[13] which is pivotal for Aβ plaques disposal. Interestingly, Aβ plaques increase mTOR's activity. However, this increase results in the down-regulation of autophagy, which prevents Aβ disposal.[13] This regulation is contradictory to the role of p53. As discussed above p53 is upregulated due to Aβ plaques. Furthermore, p53 is shown to inhibit mTOR activity,[12] which increases autophagy and, thus, Aβ plaques disposal. These contradicting results show the complicate interplay of mTOR regulators in AD.

APOE

Aβ plaques are also interacting with Apolipoprotein E (ApoE). ApoE is important in cholesterol transport and vascular pathology.[14] Furthermore, especially ApoE ε4 is a strong risk factor for AD.[15] This role of ApoE ε4 is shown from statistical and experimental evidence. Experimentally, ApoE ε4 binds to Aβ plaques, however, the role of this interaction is not yet eluded. Additionally, due to ApoE ε4's role in vascular pathology, it might also have a role in BBB integrity.[16] Lastly, ApoE ε4 is also associated with tau tangles.[17] These findings indicate the importance of ApoE in AD.

MAPT

In AD and other neurodegenerative disorders, it is believed that neuronal loss is partly caused by misfolding of tau proteins. Six different tau isomers are produced by alternative splicing of the microtubule-associated protein tau (*MAPT*) gene.[18] Tau proteins are susceptible to misfolding, which leads to tau aggregates and fibrilization.[19] This aggregation can be transmitted from neuron to neuron and induces apoptosis. Furthermore, it seems that Aβ plaques are involved in tau tangle formation since ApoE binds to tau tangles after the formation of Aβ plaques.[17] These tangles, together with Aβ plaques, ultimately lead to neuronal loss in AD.

SIRT1

Silent mating type information regulation 2 homolog 1 (SIRT1) is a family member of a group of highly conserved genes that range from genes in bacteria to humans.[20,21] This gene family is important in the ageing process, which is shown for silent information regulator 2 (Sir2) in *Saccharomyces cerevisiae*. In *S. cerevisiae*, Sir2 deleted cells had a reduction in the life span of ≈ 50% while adding a second copy of Sir2 increased the life span with ≈ 30%.[22] These results indicate the importance of the Sir gene family and human homologs were shortly thereafter found, including SIRT1.

Seven Sir human homologs are found SIRT1-7 and SIRT1 has great sequence homology with Sir2.[20] These proteins are NAD^+-dependent deacetylases and are thought to have different substrates. One role of SIRT1 is deacetylating p53, which is the primary mediator of the DNA damage response.[10] After DNA damage, the p53 is first phosphorylated and then acetylated before p53 is active and starts a cascade to induce apoptosis. However, SIRT1 prevents this cascade by deacetylating p53, which prevents cell death. Furthermore, SIRT1 also deacetylates histone H4 peptides, which might

regulate gene expression.[23] Additionally, SIRT1 levels decrease while cells age and are shown to be expressed in high levels in brain cells.[23] This indicates the probable role of SIRT1 in ageing and since AD is an age-related disease, the role of SIRT1 in AD.

In AD, cells in the brain age since they endure a rise of oxidative stress, perturbed energy homeostasis, an increase in lesions in nucleic acids and an accumulation of damaged proteins.[24] These effects result in stockpiling of damage, which triggers apoptosis. As shown above, SIRT1 already is shown to prevent apoptosis caused by p53. Furthermore, SIRT1 levels, in mouse model p25, rise due to a disruption in the calcium homeostasis, which generates oxidative stress[25]. Additionally, oxidative stress often causes DNA breaks. Due to these breaks, SIRT1 associates with the damaged DNA sites inducing repair, which promotes DNA integrity.[26] Lastly, in post-mortem human brain cells, AD patients had more Aβ plaques than the control patients.[27] In the brain regions with elevated Aβ plaques, the SIRT1 expression level was lower, which indicate that SIRT1 levels are associated with the formation of Aβ plaques. These results indicate the importance of SIRT1 in ageing and AD development (Figure 6). However, the precise mechanism is still unknown. Therefore, we tried this text-mining based procedure to find new insights in the role of SIRT1 in AD. This process resulted in three possible interesting proteins FOXQ1, S1PR1 and CLDN5.

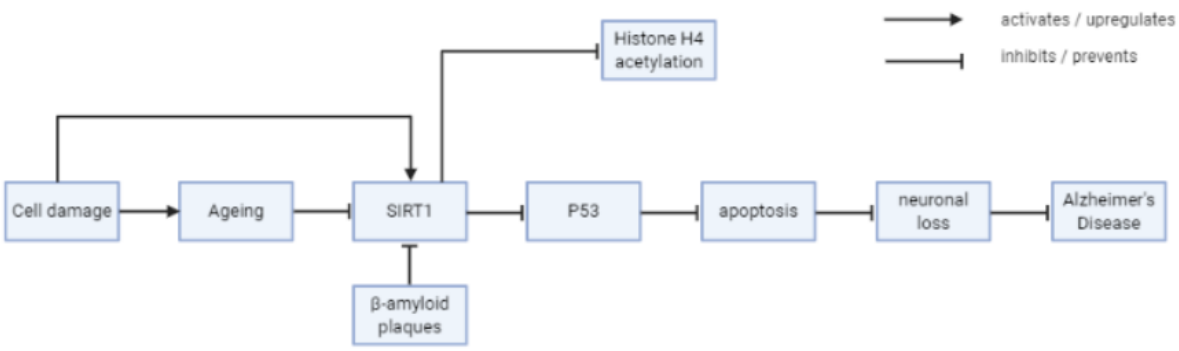


Figure 6: Schematic summary of SIRT1 in Alzheimer's Disease.

FOXQ1

One of the proteins that showed to have an interaction with SIRT1 is Forkhead box protein Q1 (FOXQ1). FOXQ1 is a member of the large Forkhead box protein family. Proteins in this family are transcription factors and, just as SIRT1, these proteins are highly conserved between different organisms.[28] FOXQ1 has many physiological functions and, especially, FOXQ1's role in cancer metastasis is widely studied.[29] Furthermore, FOXQ1 is found to be upregulated in non-small cell lung cancer tumour tissue, which also leads to an unfavourable prognosis of the patient.[30] However, FOXQ1's role in cancer development is not connected to SIRT1.

Nevertheless, SIRT1 is shown to be involved in cancer, which makes sense since cancer is also an age-related disease. Horvath e.a (2013)[31] has shown that DNA methylation (DNAm) patterns can be used to predict tissue age. Furthermore, cancer tissues have altered DNAm patterns compared to healthy tissues, which leads to altered tissue age. Some cancer types show positive age acceleration effects, while tp53 (which translates for p53) mutated cancer types exhibit negative age acceleration. Since p53 is a known SIRT1 target, perhaps affinity of SIRT1 for the mutated p53 increases, which increases deacetylation of p53. This deacetylation inhibits p53 and prevents inducing apoptosis increasing the lifetime of the cells. Additionally, overexpression of SIRT1 in ovarian cancer cells prevented apoptosis.[32] Thus, it seems that overexpression of SIRT1 in cancer prevents apoptosis while in AD SIRT1 expression is reduced, which induces more apoptosis.

Actually, FOXQ1 influences SIRT1 expression, which plays a major role in cellular senescence. FOXQ1 binds directly to the SIRT1 promoter, which upregulates SIRT1 expression.[33] This is seen in cancer cells, where FOXQ1 is often upregulated, which causes upregulation of SIRT1 and prevents cellular senescence. Furthermore, in human umbilical cord-derived stem cells (hUC-MSCs), overexpression of FOXQ1 also prevents apoptosis.[34] Additionally, transplantation of hUC-MSCs that overexpress FOXQ1 in AD mouse models improved spatial learning. In these transplanted mice models, SIRT1 expression was also significantly increased. These results reveal a major insight into the role of FOXQ1 in SIRT1 expression and AD (Figure 7).

Figure 7: Schematic summary of the interaction between FOXQ1 and SIRT1 in Alzheimer's Disease.

S1PR1

Another protein that was found that has a promising interaction is sphingosine-1-phosphate receptor 1 (S1PR1). S1PR1 binds to, as the name suggests, sphingosine-1-phosphate (S1P).[35] S1PR1 is family of the G protein-coupled receptors and is mainly localized in the plasma membrane.[36] Additionally, S1PR1 is expressed throughout the whole brain.[37] One of the major functions of S1PR1 is its essential role in vascular maturation.[38,39] Furthermore, S1PR1 plays also a role in AD.

This role in AD is not well characterised yet, however, S1P levels are shown to be important. Especially, the sphingolipid metabolism, which is disrupted in AD.[40] For instance, in AD brains, ceramidase activity is upregulated, which increases sphingosine levels. However, this rise did not result in an increase in S1P as expected, but in a decrease. Specifically, the S1P/sphingosine ratio is important in AD, which is shown to be 66% and 64% lower in the hippocampus and inferior temporal cortex, respectively.[41] Furthermore, the reduced S1P level was correlated with elevated Aβ levels and an increased amount of hyperphosphorylated tau, [40] which are two important traits in AD.

In addition, S1PR1 agonists improved cell proliferation and reduced apoptosis. Firstly, phosphorylated fingolimod acts as an agonist for S1PR1, which causes degradation of S1PR1 on lymphocytes.[42] This degeneration resulted in less Aβ plaques in the frontal cortex and hippocampus of mice models, which improved their memory. Secondly, SEW2871 acts as an agonist for S1PR1, which improved the viability of Swedish mutated APP transfected PC12 cells.[43] This enhancement was probably caused by the upregulation of SIRT1. Thus, both activation and deactivation of S1PR1 are shown to have a positive effect on cell viability, which makes it difficult to distinguish the role of S1PR1 in AD. Both agonists, indicate the importance of activation S1PR1 to enhance cell viability (Figure 8).

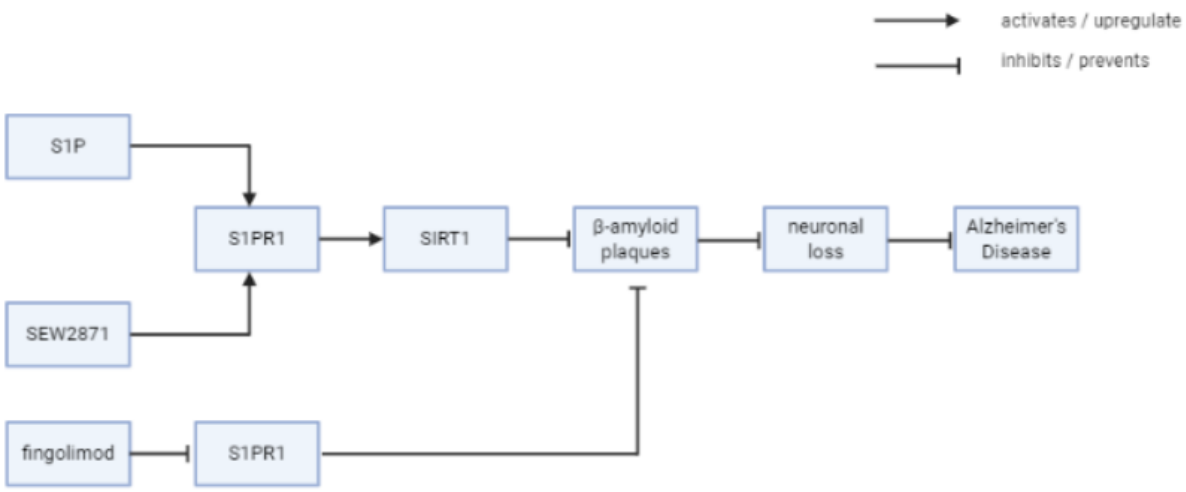


Figure 8: Schematic summary of the interaction between S1PR1 and SIRT1 in Alzheimer's Disease.

CLDN5

The last protein discussed in this review is claudin-5 (CLDN5). CLDN5 is a family member of the transmembrane proteins of which there are at least 23 members in humans.[44] CLDN5 plays a major role in tight junctions (TJs), which are important in regulating vascular permeability.[45] CLDN5 is especially found in the endothelial cells of blood cells. Furthermore, CLDN5 is also important in the breakdown of the extracellular matrix as CLDN5 activates Pro-matrix Metalloproteinase-2.[46] Notably, CLDN5 is expressed in mouse-embryo brains, which might be important in the blood-brain barrier (BBB) function.[45] Thus, CLDN5 might, via the BBB, be involved in neurodegenerative diseases, as AD.

CLDN5 protects the integrity of the BBB and regulates its permeability. In AD, removal of Aβ plaques, from the brain, is regulated by the BBB's permeability.[47] This penetrability is affected during inflammation by microglia, a type of neural cells.[48] These microglia have upregulated CLDN5 levels, which form TJs to protect the integrity of the BBB. Furthermore, in neurodegenerative patients, it was found that CLDN5 sites are methylated.[49] This methylation is linked to cognitive decline, which might be caused by downregulation of CLDN5 due to its methylation pattern. These results indicate the importance of CLDN5 in BBB's integrity, which is important in AD.

SIRT1 is capable of regulating CLDN5 by various mechanisms. Firstly, SIRT1 regulates CLDN5 by restoring the calcium homeostasis.[50] In Spinocerebellar Ataxia Type 7 (SC7, which is an inheritable neurodegenerative disorder) mouse models, SIRT1 regulates calcium homeostasis, which enables neuroprotection. Furthermore, SIRT1 overexpression in SC7 mouse models rescues the disease phenotype by promoting calcium homeostasis. This calcium homeostasis is, as discussed before, important in internalizing and down-regulating CLDN5, which disrupts the BBB permeability.[51] Secondly, SIRT1 regulates DNA methylation.[52] DNA methylation is important in macrophage, a type of white blood cell, differentiation, which is regulated by SIRT1/2. Inhibition of SIRT1/2 withdraws DNA methylation. Both SIRT1 and SIRT2 are shown to interact with DNA-methyltransferase 3 beta (DNMT3β), which causes hypermethylation of certain genes. This process might also regulate CLDN5 expression via its methylation state. This altered expression affects BBB permeability. Thus, SIRT1 regulates CLDN5 indirectly by restoring calcium homeostasis and regulating CLDN5 methylation (Figure 9).

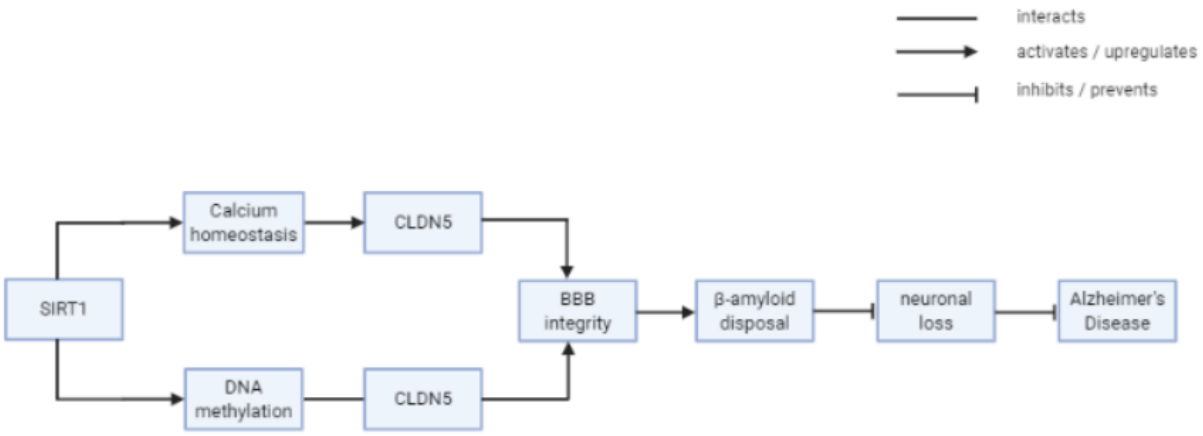


Figure 9: Schematic summary of the interaction between CLDN5 and SIRT1 in Alzheimer's Disease.

Conclusion

In this work, we were able to visualize 1305 articles, relating to Sirtuins and AD, in a protein network. Using this network, four recently discover SIRT1 regulators in AD were identified. However, one of the proteins, PRKAA2, has a synonym, AMPK, of which the connection with SIRT1 is quite well described already. Furthermore, one key protein, APP, was missing from the network. These results indicate that analysing networks based on text-mining should be done cautiously. Although, the three other proteins: FOXQ1, S1PR1 and CLDN5 showed interesting interactions with SIRT1.

In cancer cells, FOXQ1 is upregulated, which increases SIRT1 expression. This rise of SIRT1 prevents apoptosis. Furthermore, AD mouse models were transplanted with umbilical cord-derived stem cells that were overexpressing FOXQ1. This transplantation improved spatial learning in these AD mice, potentially by upregulating SIRT1 expression.

Furthermore, agonists of S1PR1 also reduce apoptosis. This reduction might also be caused by upregulation of SIRT1. However, degradation of S1PR1 is also shown to benefit AD mouse models due to lowering levels of Aβ plaques. These contradicting results make it difficult to establish the interplay of SIRT1 and S1PR1 in AD. However, there is definitely a connection between the two proteins.

Lastly, SIRT1 regulates CLDN5 indirectly, which might improve Aβ deposition. Aβ removal is facilitated by the BBB. This BBB consists of CLDN5, which is important in the correct functioning of this BBB. CLDN5 is possibly regulated by SIRT1 in two indirect manners, which are restoring calcium homeostasis and DNA methylation. If the calcium homeostasis is disrupted, CLDN5 expression is also disrupted, which has an impact on BBB integrity. Furthermore, in AD patients, the CLDN5 gene is methylated, which also influences the expression. Both mechanisms are regulated by SIRT1 expression. However, the direct connection between SIRT1 and CLDN5 has not been found yet.

These three interesting interactions with SIRT1 show the strength and possibility of this novel text-mining based approach. Furthermore, the workflow in the Jupiter Notebook automatically retrieves the information from the literature, which makes it reproducible and sharable.

Future Outlook

Although we successfully created a reusable workflow that extracts meaningful information, there is still room for improvements. These two improvements are based on the two difficulties that arose in the results of this work.

Firstly, we found an interesting, recently discovered interaction of SIRT1, PRKAA2. However, PRKAA2 is a synonym for AMPK, which was also found in the protein network but with an average publication year of 2015. To circumvent this problem, for each found protein, the synonym can be retrieved from the HGNC. Then, the results of the synonyms can be added together which prevents this problem.

Secondly, the weight scoring equation can be improved to find more meaningful interactions. An often-used score for information retrieval is the term frequency-inverse document frequency (TFIDF). The TFIDF value is based on the number of times a word appears in the document. Additionally, two proteins that are found in the same sentence often have a closer relationship with each other. Thus, this closeness can also improve weight scoring. However, in text-mining, it is still hard to distil the sentiment of a sentence. So, improvements can be made in text-mining and more specifically, in weight scoring.

Lastly, text-mining might be used to search for new drug candidates. These candidates might be found using the workflow of Tshitoyan e.a. (2019)[2], in which they recommended materials for functional applications. It would be interesting to see if this method would also yield new promising drug candidates.

References

1. Brookmeyer, R., Johnson, E., Ziegler-Graham, K. & Arrighi, H. M. Forecasting the global burden of Alzheimer's disease. *Alzheimer's & Dementia* **3**, 186–191 (2007).
2. Tshitoyan, V. *et al.* Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
3. Palmblad, M. Visual and Semantic Enrichment of Analytical Chemistry Literature Searches by Combining Text Mining and Computational Chemistry. *Anal. Chem.* **91**, 4312–4316 (2019).
4. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **47**, D607–D613 (2019).
5. Guler, A. T., Waaijer, C. J. F., Mohammed, Y. & Palmblad, M. Automating bibliometric analyses using Taverna scientific workflows: A tutorial on integrating Web Services. *Journal of Informetrics* **10**, 830–841 (2016).
6. Loo, D. T. *et al.* Apoptosis is induced by beta-amyloid in cultured central nervous system neurons. *Proceedings of the National Academy of Sciences* **90**, 7951–7955 (1993).
7. Behl, C. Hydrogen peroxide mediates amyloid β protein toxicity. *Cell* **77**, 817–827 (1994).
8. Kerpedjiev, P. The 20 Most Studied Genes. *Empty Pipes* http://emptypipes.org/2014/12/08/gene-popularity/ (2014).
9. Lin, Y., Ma, W. & Benchimol, S. Pidd, a new death-domain–containing protein, is induced by p53 and promotes apoptosis. *Nat Genet* **26**, 122–127 (2000).
10. Vaziri, H. *et al.* hSIR2SIRT1 Functions as an NAD-Dependent p53 Deacetylase. *Cell* **107**, 149–159 (2001).
11. Kitamura, Y. *et al.* Changes of p53 in the Brains of Patients with Alzheimer's Disease. *Biochemical and Biophysical Research Communications* **232**, 418–421 (1997).
12. Feng, Z., Zhang, H., Levine, A. J. & Jin, S. The coordinate regulation of the p53 and mTOR pathways in cells. *Proceedings of the National Academy of Sciences* **102**, 8204–8209 (2005).

13. Caccamo, A., Majumder, S., Richardson, A., Strong, R. & Oddo, S. Molecular Interplay between Mammalian Target of Rapamycin (mTOR), Amyloid-β, and Tau: EFFECTS ON COGNITIVE IMPAIRMENTS. *J. Biol. Chem.* **285**, 13107–13120 (2010).
14. Strittmatter, W. J. *et al.* Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proceedings of the National Academy of Sciences* **90**, 1977–1981 (1993).
15. Saunders, A. M. *et al.* Association of apolipoprotein E allele 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* **43**, 1467–1467 (1993).
16. Riphagen, J. M. *et al.* Linking APOE-ε4, blood-brain barrier dysfunction, and inflammation to Alzheimer's pathology. *Neurobiology of Aging* **85**, 96–103 (2020).
17. Farfel, J. M., Yu, L., De Jager, P. L., Schneider, J. A. & Bennett, D. A. Association of APOE with tau-tangle pathology with and without β-amyloid. *Neurobiology of Aging* **37**, 19–25 (2016).
18. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
19. de Calignon, A. *et al.* Propagation of Tau Pathology in a Model of Early Alzheimer's Disease. *Neuron* **73**, 685–697 (2012).
20. Frye, R. A. Characterization of Five Human cDNAs with Homology to the Yeast SIR2 Gene: Sir2-like Proteins (Sirtuins) Metabolize NAD and May Have Protein ADP-Ribosyltransferase Activity. *Biochemical and Biophysical Research Communications* **260**, 273–279 (1999).
21. Frye, R. A. Phylogenetic Classification of Prokaryotic and Eukaryotic Sir2-like Proteins. *Biochemical and Biophysical Research Communications* **273**, 793–798 (2000).
22. Kaeberlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. *Genes & Development* **13**, 2570–2580 (1999).

23. Michishita, E., Park, J. Y., Burneskis, J. M., Barrett, J. C. & Horikawa, I. Evolutionarily Conserved and Nonconserved Cellular Localizations and Functions of Human SIRT Proteins. *MBoC* **16**, 4623–4635 (2005).
24. Mattson, M. P. & Magnus, T. Ageing and neuronal vulnerability. *Nat Rev Neurosci* **7**, 278–294 (2006).
25. Kim, D. *et al.* SIRT1 deacetylase protects against neurodegeneration in models for Alzheimer's disease and amyotrophic lateral sclerosis. *EMBO J* **26**, 3169–3179 (2007).
26. Oberdoerffer, P. *et al.* SIRT1 redistribution on chromatin promotes genomic stability but alters gene expression during aging. *Cell* **135**, 907–918 (2008).
27. Cao, K. *et al.* Reduced expression of SIRT1 and SOD-1 and the correlation between these levels in various regions of the brains of patients with Alzheimer's disease. *J Clin Pathol* **71**, 1090–1099 (2018).
28. Kaestner, K. H., Knochel, W. & Martinez, D. E. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev.* **14**, 142–146 (2000).
29. Li, Y. *et al.* Forkhead box Q1: A key player in the pathogenesis of tumors (Review). *International Journal of Oncology* **49**, 51–58 (2016).
30. Li, L. *et al.* Potentiality of forkhead box Q1 as a biomarker for monitoring tumor features and predicting prognosis in non-small cell lung cancer. *J Clin Lab Anal* **34**, (2020).
31. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biology* **14**, R115 (2013).
32. Xia, X. Y. *et al.* MicroRNA-506-3p inhibits proliferation and promotes apoptosis in ovarian cancer cell via targeting SIRT1/AKT/FOXO3a signaling pathway. *neo* (2020) doi:10.4149/neo_2020_190517N441.
33. Wang, P. *et al.* FOXQ1 regulates senescence-associated inflammation via activation of SIRT1 expression. *Cell Death Dis* **8**, e2946–e2946 (2017).

34. Zhang, T. *et al.* Overexpression of FOXQ1 enhances anti-senescence and migration effects of human umbilical cord mesenchymal stem cells in vitro and in vivo. *Cell Tissue Res* **373**, 379–393 (2018).
35. Lee, M. Sphingosine-1-Phosphate as a Ligand for the G Protein-Coupled Receptor EDG-1. *Science* **279**, 1552–1555 (1998).
36. Sanchez, T. & Hla, T. Structural and functional characteristics of S1P receptors. *J. Cell. Biochem.* **92**, 913–922 (2004).
37. Chae, S.-S., Proia, R. L. & Hla, T. Constitutive expression of the S1P1 receptor in adult tissues. *Prostaglandins & Other Lipid Mediators* **73**, 141–150 (2004).
38. Hla, T. & Maciag, T. An abundant transcript induced in differentiating human endothelial cells encodes a polypeptide with structural similarities to G-protein-coupled receptors. *J. Biol. Chem.* **265**, 9308–9313 (1990).
39. Liu, Y. *et al.* Edg-1, the G protein–coupled receptor for sphingosine-1-phosphate, is essential for vascular maturation. *J. Clin. Invest.* **106**, 951–961 (2000).
40. He, X., Huang, Y., Li, B., Gong, C.-X. & Schuchman, E. H. Deregulation of sphingolipid metabolism in Alzheimer's disease. *Neurobiology of Aging* **31**, 398–408 (2010).
41. Couttas, T. A. *et al.* Loss of the neuroprotective factor Sphingosine 1-phosphate early in Alzheimer's disease pathogenesis. *acta neuropathol commun* **2**, 9 (2014).
42. Carreras, I. *et al.* Dual dose-dependent effects of fingolimod in a mouse model of Alzheimer's disease. *Sci Rep* **9**, 10972 (2019).
43. Czubowicz, K., Wójtowicz, S., Wencel, P. L. & Strosznajder, R. P. The role of ceramide and SEW 2871 in the transcription of enzymes involved in amyloid b precursor protein metabolism in an experimental model of Alzheimer's disease. *fn* **56**, 196–205 (2018).
44. Liu, F. *et al.* Systems Proteomics View of the Endogenous Human Claudin Protein Family. *J. Proteome Res.* **15**, 339–359 (2016).

45. Morita, K., Sasaki, H., Furuse, M. & Tsukita, S. Endothelial claudin: claudin-5/TMVCF constitutes tight junction strands in endothelial cells. *The Journal of Cell Biology* **147**, 185–194 (1999).
46. Miyamori, H. *et al.* Claudin Promotes Activation of Pro-matrix Metalloproteinase-2 Mediated by Membrane-type Matrix Metalloproteinases. *J. Biol. Chem.* **276**, 28204–28211 (2001).
47. Cirrito, J. R. P-glycoprotein deficiency at the blood-brain barrier increases amyloid- deposition in an Alzheimer disease mouse model. *Journal of Clinical Investigation* **115**, 3285–3290 (2005).
48. Haruwaka, K. *et al.* Dual microglia effects on blood brain barrier permeability induced by systemic inflammation. *Nat Commun* **10**, 5816 (2019).
49. Hüls, A. *et al. Brain DNA Methylation Patterns in* CLDN5 *Associated With Cognitive Decline*. http://biorxiv.org/lookup/doi/10.1101/857953 (2019) doi:10.1101/857953.
50. Stoyas, C. A. *et al.* Nicotinamide Pathway-Dependent Sirt1 Activation Restores Calcium Homeostasis to Achieve Neuroprotection in Spinocerebellar Ataxia Type 7. *Neuron* S0896627319309833 (2019) doi:10.1016/j.neuron.2019.11.019.
51. Nokkari, A. *et al.* Implication of the Kallikrein-Kinin system in neurological disorders: Quest for potential biomarkers and mechanisms. *Progress in Neurobiology* **165–167**, 26–50 (2018).
52. Li, T. *et al.* SIRT1/2 orchestrate acquisition of DNA methylation and loss of histone H3 activating marks to prevent premature activation of inflammatory genes in macrophages. *Nucleic Acids Research* **48**, 665–681 (2020).