creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

이학석사 학위논문

Bioinformatic analysis on pathological association of human aminoacyl-tRNA synthetases and their protein network with neurological diseases

인간 단백질합성효소들에 의한 네트워크와 신경질환과의 병리적 연관성에 관한 생물정보학적 분석

2017 년 2 월

서울대학교 융합과학기술대학원

분자의학 및 바이오제약학과 의약생명과학전공

한동진

Bioinformatic analysis on pathological association of human aminoacyl-tRNA synthetases and their protein network with neurological diseases

인간 단백질합성효소들에 의한 네트워크와 신경질환과의 병리적 연관성에 관한 생물정보학적 분석

지도교수 김 성 훈
이 논문을 이학석사 학위논문으로 제출함

2017 년 2 월

서울대학교 융합과학기술대학원
분자의학 및 바이오제약학과 의약생명과학전공
한동진

한동진의 이학석사 학위논문을 인준함
2017 년 2 월

위 원 장 _____ (인)

부 위 원 장 _____ (인)

위 원 _____ (인)

Abstract

Bioinformatic analysis on pathological association of human aminoacyl-tRNA synthetases and their protein network with neurological diseases

Dong-Jin Han

Department of Molecular Medicine and Biopharmaceutical Science,

The Graduate School of Convergence Science and Technology,

Seoul National University

Aminoacyl-tRNA synthetases (ARSs) are essential enzymes ligating specific amino acids to their cognate tRNA for protein biosynthesis. It is reported that they are also involved in many signaling pathways as crucial mediators. In these multiple activities, they are associated with various human diseases. Many mutations of ARSs have been found to be associated with various neurological diseases. Here, we systematically investigated the statistical association of ARSs and their associated factors such as ARS-interacting multi-functional proteins (AIMPs) with various neurological diseases. A total network of ARSs/AIMPs and their

interacting factors was constructed using three protein-protein interaction (PPI) databases. In this network, 586 factors were identified as first-neighbors that are suggested to be directly linked to ARSs/AIMPs and 13,539 factors were identified as second-neighbors that are indirectly linked to ARSs/AIMPs via the first-neighbors. Among the first- and second-neighbors, we selected 1,772 genes associated with 27 neurological diseases (neurological-disease-associated genes; NAGs) from public databases and the literature and identified 86 and 687 factors as the first- and second-neighbor NAGs, respectively.

We retrieved 67 gene expression datasets of 24 neurological diseases from Gene Expression Omnibus (GEO) and ArrayExpress. The gene expression profiles of ARSs/AIMPs and their neighboring NAGs were compared with negative controls (non-NAGs). We obtained *P* values for each dataset and combined them for each disease. Then, the *P* values for each disease were combined into a value representing the whole set of 24 neurological diseases. Additionally, we created a subnetwork representing biological processes and *P* values using the Database for Annotation, Visualization, and Integrated Discovery (DAVID).

Quite a few ARSs/AIMPs and first- and second-neighbor NAGs were differentially expressed genes (DEGs) in neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease. In summary, 20 human cytosolic ARSs and 3 AIMPs are strongly connected to diverse NAGs and are

differentially expressed in neurological diseases, indicating their implication in these diseases.

Keywords: ARS, Neurological disease, mRNA expression profile, DEG, PPI, Network, Gene ontology biological process, NAG

Student ID: 2014-24873

Contents

List of figures and tables

List of abbreviations

AD: Alzheimer's disease

ARSs/AIMPs: Aminoacyl-tRNA synthetases and ARS-interacting multifunctional proteins

DAVID: Database for annotation, visualization and integrated discovery

DEGs: Differentially expressed genes

GEO: Gene expression omnibus

GOBP: Gene ontology biological process

NAGs: Neurological-disease-associated genes

PPI: Protein-protein interaction

Introduction

Aminoacyl-tRNA synthetases (ARSs) are enzymes necessary for aminoacylation of tRNA and their catalytic activities are indispensable for protein biosynthesis. In particular, mammalian ARSs form the multi-tRNA synthetase complex (MSC) with three ARS-interacting multifunctional proteins (AIMPs)[1,2]; this complex serves as a molecular repository for the control of diverse signaling pathways.[3,4]

Many diseases including neurological and immune-system-related diseases and cancers are associated with ARSs/AIMPs.[3] We have previously suggested the association of ARSs/AIMPs with various types of cancer through the systematic analysis of their expression, copy number variations, and mutations in conjunction with their interacting factors.[5,6] Nonetheless, systematic investigation of the potential implication of ARSs/AIMPs in neurological diseases and gene expression profiling studies on both ARSs and AIMPs have not yet been conducted. Furthermore, many cytosolic ARS and AIMP mutations have been found in various neurological diseases (Table 1). Although some of the mutations are associated with reduced catalytic activity of ARSs, other mutations are not related to catalytic activity.[7] Therefore, non-canonical functions (not related to protein biosynthesis) of ARSs can be implicated in neurological diseases. Here, we tested whether

ARSs/AIMPs are implicated in neurological diseases via aberrant expression or interactions with neurological-disease-associated factors.

Several experimental results have shown these possibilities. For instance, *AIMP2* overexpressing transgenic mice show a loss of dopaminergic neurons: the major cause of Parkinson's disease.[8] Moreover, mutant forms of soluble superoxide dismutase 1 (*SOD1*), found in amyotrophic lateral sclerosis (ALS) patients,[9] interact with lysyl-tRNA synthetase (*KARS*), implying its potential association with ALS.[3] In a study on metabolites collected from the plasma of patients with Alzheimer's disease (AD), metabolomic profiles that were associated with the aminoacyl-tRNA biosynthesis pathway were significantly different between the disease group and the cognitively healthy group.[10] Moreover, impairment of the canonical function of ARSs by oxidative stress can contribute to mistranslation.[11,12] If *APP* or tau protein is mistranslated, these proteins can be misfolded due to a change of amino acid sequence. In the end, misfolded Aβ or tau proteins can be propagated by a mechanism similar to that of prions in Creutzfeldt-Jakob disease.[13]

In this study, we examined mRNA expression profiles of 20 cytoplasmic ARSs and 3 AIMPs, and their neighboring factors known to be associated with 24 neurological diseases, including both peripheral nervous system diseases and central nervous system diseases. Additionally, we created a hypothetical network of differentially expressed genes (DEGs) and gene ontology biological

processes (GOBPs).

Methods

The source of NAGs and non-NAGs

Because we needed genes both linked to ARSs/AIMPs and associated with a neurological disease, we collected neurological disease-associated genes (NAGs) and protein-protein interaction (PPI) data. Before selecting the NAGs, we needed to consider which neurological diseases are most related to ARSs/AIMPs. We selected 27 neurological diseases based on the prevalence of diseases and review articles.[3,14] We then used four databases, PharmDB[15], DisGeNET[16], ClinVar[17], and NECTAR[18], to select the genes that are likely to be involved in these neurological diseases. Additionally, the genes that were not included in these four databases were retrieved from the literature.[9,19-68] The genes involved in at least one of the 27 diseases were labeled as NAGs. BioGRID[69], intAct[70], and PharmDB PPI data were used to classify NAGs into the first-neighbor NAGs (direct interactors of ARSs/AIMPs) and second-neighbor NAGs (direct interactors of the first neighbors). Among the total of 1,772 NAGs, 86 and 687 were identified as the first- and second-neighbor NAGs, respectively. Non-NAGs were selected from PharmDB genes that are not associated with any diseases or drug actions. From the remaining set, all genes in DisGeNET and all NAGs were excluded, and thus

the final number of non-NAGs was 881.

Normalization of gene expression datasets and calculation of a combined *P* value

A total of 67 mRNA expression datasets were retrieved from Gene Expression Omnibus (GEO)[71] and ArrayExpress[72]. We normalized 59 Affymetrix microarray gene expression datasets using the GC-RMA normalization method[73] and PLIER algorithm[74]. The other four Agilent platforms, three Illumina platforms, and one custom microarray platform were normalized using the quantile normalization method[75]. Each dataset was normalized as previously described.[6] In brief, all the genes were subdivided into two groups, expressed group and non-expressed group, based on a specific threshold intensity value. Student's *t*-test, the Wilcoxon rank-sum test, and the $\log_2$-median-ratio test were applied to genes included in the expressed group. In the $\log_2$-median-ratio test, random permutations of samples were needed to create empirical distributions of the hypothesis. Then, the empirical distributions were used to calculate the *P* value for each gene. Finally, the *P* values from the three tests were combined by Stouffer's method[76]. Again, we combined the *P* values calculated from different datasets using Stouffer's method. Therefore, the *P* values for each gene of the 67 datasets for the 24 neurological diseases were combined into a single *P* value.

Orthologous gene mapping

The organisms of the gene expression datasets used for this study were humans, mice, rats, and gray mouse lemurs. Nevertheless, a human microarray chip was used for the gray mouse lemur dataset. Therefore, we needed to change the mouse or rat Entrez gene IDs to their corresponding human Entrez gene IDs using the Mouse Genome Database (MGD)[77].

Gene ontology analysis by means of the DAVID functional annotation tool

A hypothetical subnetwork of AD was created by means of the DAVID functional annotation tool. All the input genes were DEGs of at least one AD dataset of the five, and all of them were first- or second-neighbor NAGs. After running DAVID, we selected an appropriate gene ontology biological process (GOBP) term whose false discovery rate (FDR) was less than 0.05. If a gene had two or more GOBP terms, we selected the most appropriate term for the gene. Finally, 108 genes were selected, each of which had a single GOBP term.

Network and heat map visualization and calculation of average shortest paths

Cytoscape 3.3.0[78] was used for constructing the whole network

including ARSs/AIMPs with the first- and second-neighbors (Figure 6) and a hypothetical ARSs/AIMPs subnetwork of AD (Figure 9B). In addition, MATLAB R2008b was employed for heat map visualization. We generated Figure 8A using the *imagesc* function. The heat map of ARSs/AIMPs mRNA expression (Figure 8B) was generated using the *clustergram* function and the complete linkage cluster algorithm. The heat map of ARSs/AIMPs biological processes (Figure 9C) was created by the *pcolor* function.

To calculate an average shortest path, we set up a symmetric n x n sparse matrix s. If nodes *i* and *j* are connected directly, s_{ij} equals one. If not, s_{ij} equals zero. We filled out the sparse matrix and calculated the shortest path of every node to the other nodes using MATLAB. Because we calculated the shortest path between each node involved in 13 biological processes and each ARS, we could calculate means of the shortest path between nodes involved in the same biological process and each ARS (Figure 9C).

Results

ARSs are related to neurological diseases

Because there were too many neurological diseases, we selected the diseases that are known for their relation to ARSs/AIMPs.[3,14] Besides, we included AD, bipolar disease, and schizophrenia for the disease list although there is no clear evidence for the involvement of ARSs/AIMPs in these diseases. It is important to note that the disease list used for the selection of NAGs differs from the disease list used for the gene expression dataset collection (Table 2) because there are many diseases not present in GEO or ArrayExpress. Instead, the disease list used for the gene expression dataset was compiled based on five papers[79–83] on central nervous system diseases.

In the present study, we show that ARSs/AIMPs are associated with neurological diseases when viewed from four perspectives (Figure 1). Gene expression datasets were used for calculating *P* values, and DEGs were selected using these *P* values. Not only ARSs but also their neighbor genes can play a role in the pathophysiology or pathogenesis of neurological diseases. Therefore, we created an ARS/AIMP network using public PPI data. In addition, we collected NAGs to see the expression patterns of genes known for their association with neurological diseases. Lastly,

gene ontology (GO) analysis was performed using DAVID, because we needed to know which genes are involved in a specific biological process. All these results support the notion that ARSs/AIMPs and their neighbor genes are differentially expressed in neurological diseases, specifically neurodegenerative diseases (Figures 8 and 9).

The combined *P* value of each gene in 24 neurological diseases

After selecting the neurological diseases for the gene expression dataset collection, we retrieved the gene expression data from GEO and ArrayExpress. Although we compiled a limited disease list, there were still too many datasets in these two databases. Therefore, we needed to narrow the datasets down by the following criteria: (i) samples in the dataset should be subdivided into healthy controls and disease groups; (ii) only *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus* can be included in the datasets (Figure 2). As an exception, GSE21779, which used the gray mouse lemur as an experimental organism, was included in our datasets. Each dataset was normalized as specified in Table 3. To identify which gene was differentially expressed, we calculated the *P* value of each probe. Student's *t*-test, the Wilcoxon rank-sum test, and the $\log_2$-median-ratio test were performed to obtain a combined *P* value of these three tests (Figure 3). Because multiple probes corresponded to a single gene in some cases, we chose the probe with the

smallest P value. Each dataset had different numbers of samples and probes. In addition, some P values were not available, because some datasets did not have the probe for a specific gene. Therefore, we used only P values available in at least 10 datasets. We wanted to know the general expression profiles of each of the 24 neurological diseases and the combined profile for all. Therefore, we combined the P values of the same disease for each gene. Then, the P values of each disease were combined into a single P value that represented the significance of a gene in all 24 neurological diseases (Figure 4).

Construction of a network of ARSs/AIMPs and their neighbor NAGs

Because ARSs/AIMPs may be involved in neurological diseases directly or indirectly (via their neighboring factors), we retrieved PPI data from BioGRID, intAct, and PharmDB. These PPI data were merged into a total of 229,765 PPIs (Figure 5). Using these PPI data, we constructed the network of 23 ARSs/AIMPs with 586 first neighbors, and 13,539 second neighbors (Figure 6, red, yellow and black circles, respectively). We selected a limited number of genes associated with neurological diseases according to public databases and the literature (Figure 7A). The numbers of genes involved in each disease are summarized in Table 4. The non-NAGs were obtained by removing the genes associated with drugs or diseases

from the total set of PharmDB genes (Figure 7B).

Differentially expressed ARSs and their neighbor NAGs were the most numerous in Alzheimer's disease

We surveyed the proportions of DEGs of four groups (ARSs/AIMPs, first- and second-neighbor NAGs, and non-NAGs) in 24 neurological diseases (Figure 8A). As expected, the DEG proportions of ARSs/AIMPs and first- and second-neighbor NAGs were higher than those of non-NAGs in the 24 neurological diseases. The DEG proportions of the three groups were particularly high in neurodegenerative diseases such as AD and Parkinson's disease. The *P* values of each gene included in the ARS/AIMP group are presented as a heat map (Figure 8B). It is noteworthy that all the ARSs/AIMPs were differentially expressed (*P* value < 0.05) in AD except for glutaminyl- (*QARS*), threonyl- (*TARS*), and cysteinyl-tRNA synthetase (*CARS*). Moreover, *AIMP2*, which has a pathological connection with Parkinson's disease,[3] was differentially expressed in Parkinson's disease.

A hypothetical subnetwork shows proximity between ARSs/AIMPs and AD-related genes

Because the DEG proportion of the ARS/AIMP group was the highest in AD among the 24 neurological diseases, we wanted to

find out which biological processes are involved both in ARSs/AIMPs neighbor genes and in AD. Therefore, we used the DAVID functional annotation tool with 379 input genes composed of first- and second-neighbor NAGs. In addition, these 379 input genes are DEGs in at least one AD dataset among the five datasets (Figure 9A). Given that protein interactions can be depicted as a network,[84] the result of the DAVID analysis can be presented as a hypothetical subnetwork (Figure 9B). We calculated a *P* value for each of the 24 neurological diseases; thus, we used node size to represent the AD *P* value. Colors of each node mean the number of datasets. For example, glutamyl-prolyl-tRNA synthetase (*EPRS*) is downregulated in GSE1297 and at the same time upregulated in GSE5281. As a result, *EPRS* has a zero-fold change because one minus one equals zero.

Many features of a network can be quantified, and one of them is the shortest path between two nodes.[84] We prioritized these 13 biological processes on the basis of an average shortest path between ARSs/AIMPs and genes involved in each biological process. Among the 13 biological processes, the "MAPK cascade" and "Response to oxidative stress" were the nearest to ARSs/AIMPs except for "tRNA aminoacylation" according to the average shortest path (Figure 9C). Consequently, we needed to verify whether ARSs/AIMPs are related to these two biological processes and whether the relation is associated with AD.

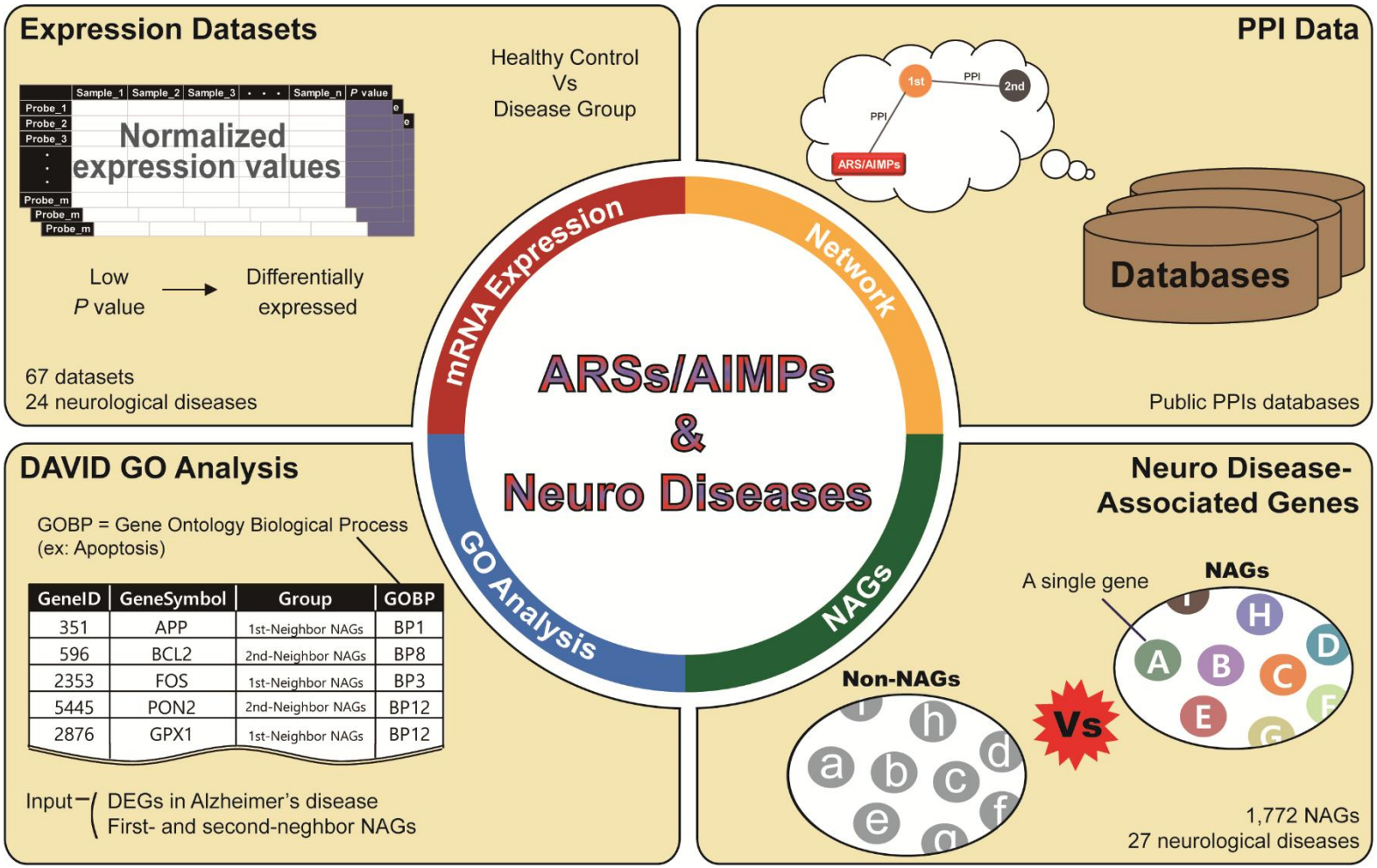

Expression Datasets
Healthy Control
Vs
Disease Group
Sample_1
Sample_2
Sample_3
Sample_n
P value
Probe_1
Probe_2
Probe_3
Probe_m
Normalized expression values
Low P value
Differentially expressed
67 datasets
24 neurological diseases
PPI Data
1st
2nd
PPI
PPI
ARS/AIMPs
Databases
Public PPIs databases
mRNA Expression
Network
GO Analysis
NAGs
ARSs/AIMPs & Neuro Diseases
DAVID GO Analysis
GOBP = Gene Ontology Biological Process
(ex: Apoptosis)
GeneID | GeneSymbol | Group | GOBP
351 | APP | 1st-Neighbor NAGs | BP1
596 | BCL2 | 2nd-Neighbor NAGs | BP8
2353 | FOS | 1st-Neighbor NAGs | BP3
5445 | PON2 | 2nd-Neighbor NAGs | BP12
2876 | GPX1 | 1st-Neighbor NAGs | BP12
Input
DEGs in Alzheimer's disease
First- and second-neghbor NAGs
Neuro Disease-Associated Genes
A single gene
NAGs
Non-NAGs
Vs
1,772 NAGs
27 neurological diseases

Figure 1. Four perspectives on the relation between ARSs/AIMPs and neurological diseases. This figure shows how we analyzed the relation between ARSs/AIMPs and neurological diseases. DEGs were selected based on the combined *P* values from multiple datasets. PPI data were used for constructing an ARS/AIMP network consisting of first- and second-neighbors. The limited number of genes known for their relations with neurological diseases were collected from public databases and the literature. Finally, we conducted a DAVID GO analysis to identify biological processes related to these genes.

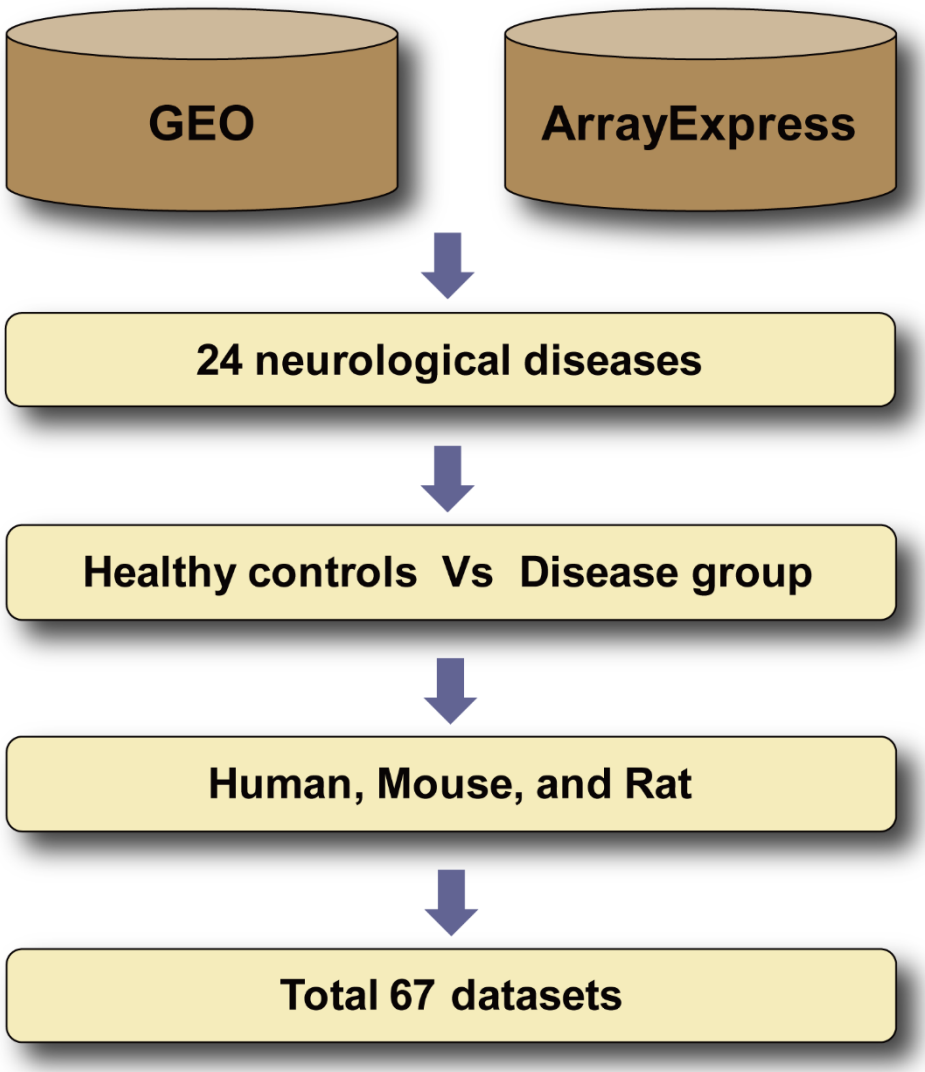


Figure 2. Selection criteria for gene expression datasets. First, 24 neurological diseases associated with gene expression datasets were found in GEO and ArrayExpress. Then, we chose datasets with both healthy controls and disease groups. In addition, all the mRNA expression datasets used for this study included only three organisms (except for one dataset): *H. sapiens*, *M. musculus*, and *R. norvegicus*. Finally, 67 datasets were selected.

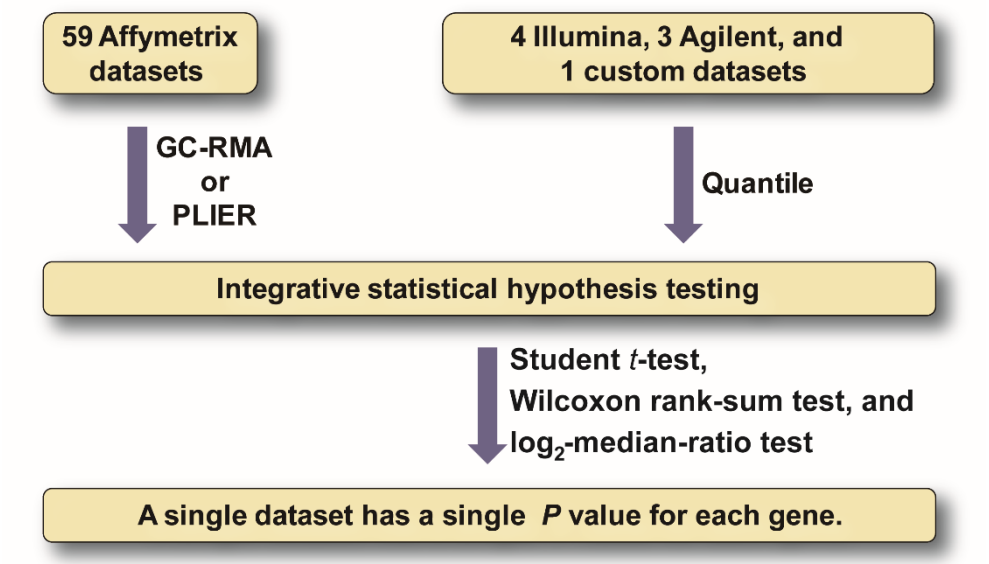


Figure 3. The data preprocessing method. We normalized 59 Affymetrix platform microarray data using the GC-RMA or PLIER algorithm. The other eight gene expression datasets were normalized by the quantile normalization method. Student's *t*-test, the Wilcoxon rank-sum test, and the $\log_2$-median-ratio test were carried out. After that, *P* values from these three tests were combined by Stouffer's method. Therefore, a single dataset had a single *P* value for each gene.

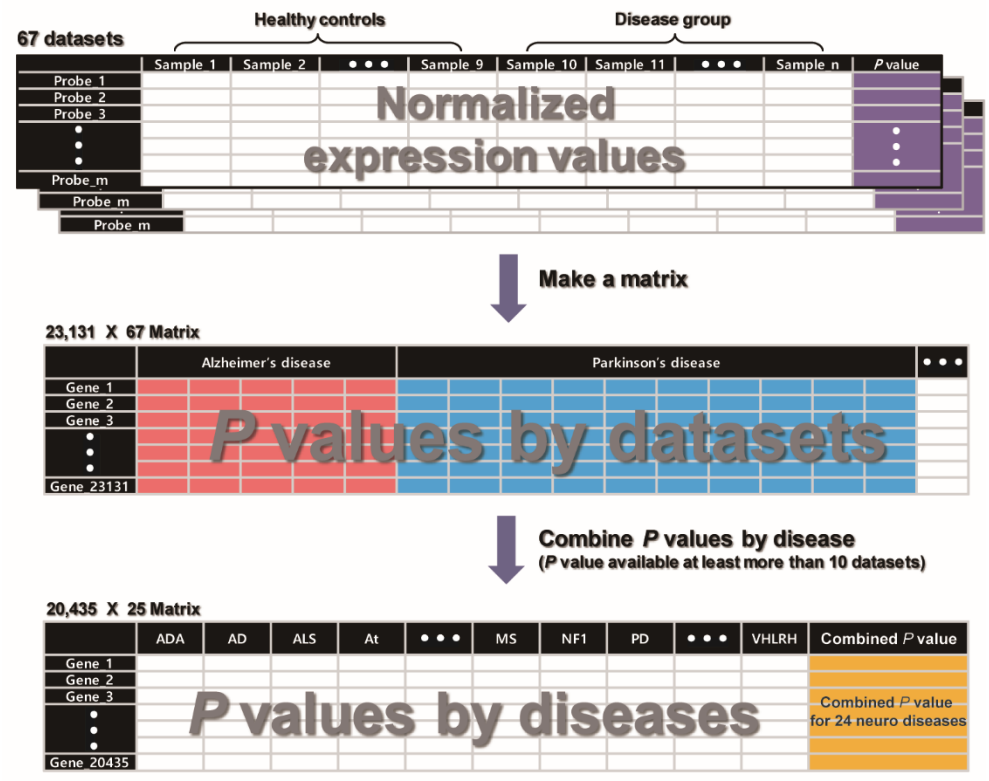


Figure 4. Combining *P* values by Stouffer's method. *P* values of dataset for the same disease (purple color) were combined by Stouffer's method. For example, five *P* values of Gene_1 were combined in AD (red color). Thus, a gene had a single *P* value in a disease. Once again, we combined all the *P* values of the 24 neurological diseases by Stouffer's method (orange color).

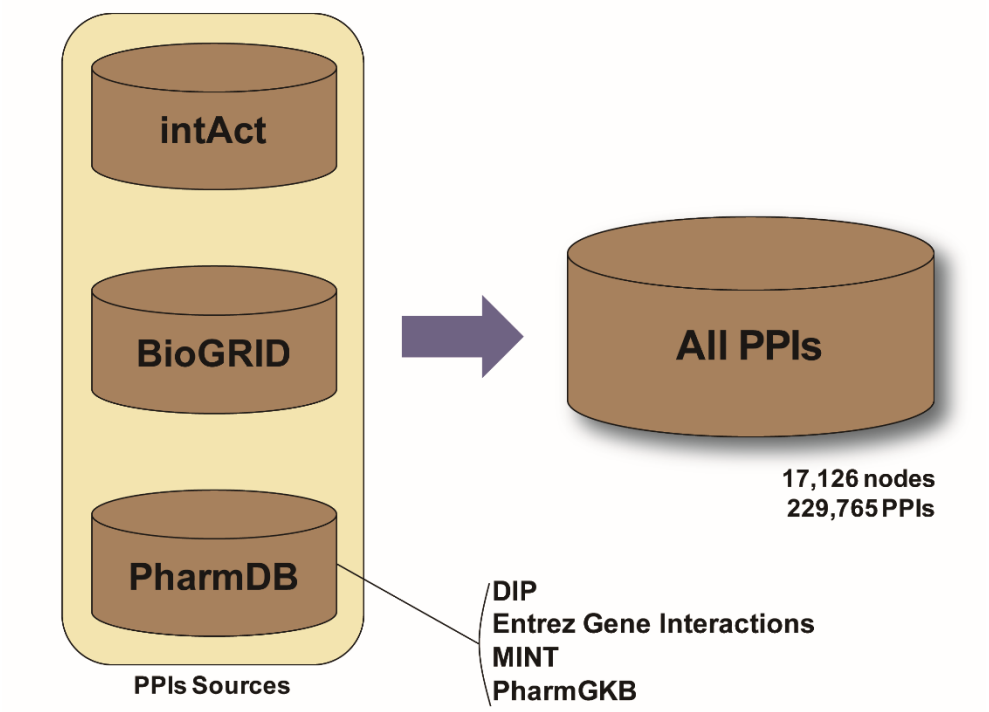


Figure 5. Collection of the protein-protein interaction (PPI) data. All the PPI data were retrieved from three PPI databases: intAct, BioGRID, and PharmDB. PharmDB PPI data came from DIP, Entrez Gene Interactions, MINT, and PharmGKB PPI data. The total PPI data were composed of 17,126 nodes and 229,765 PPIs.

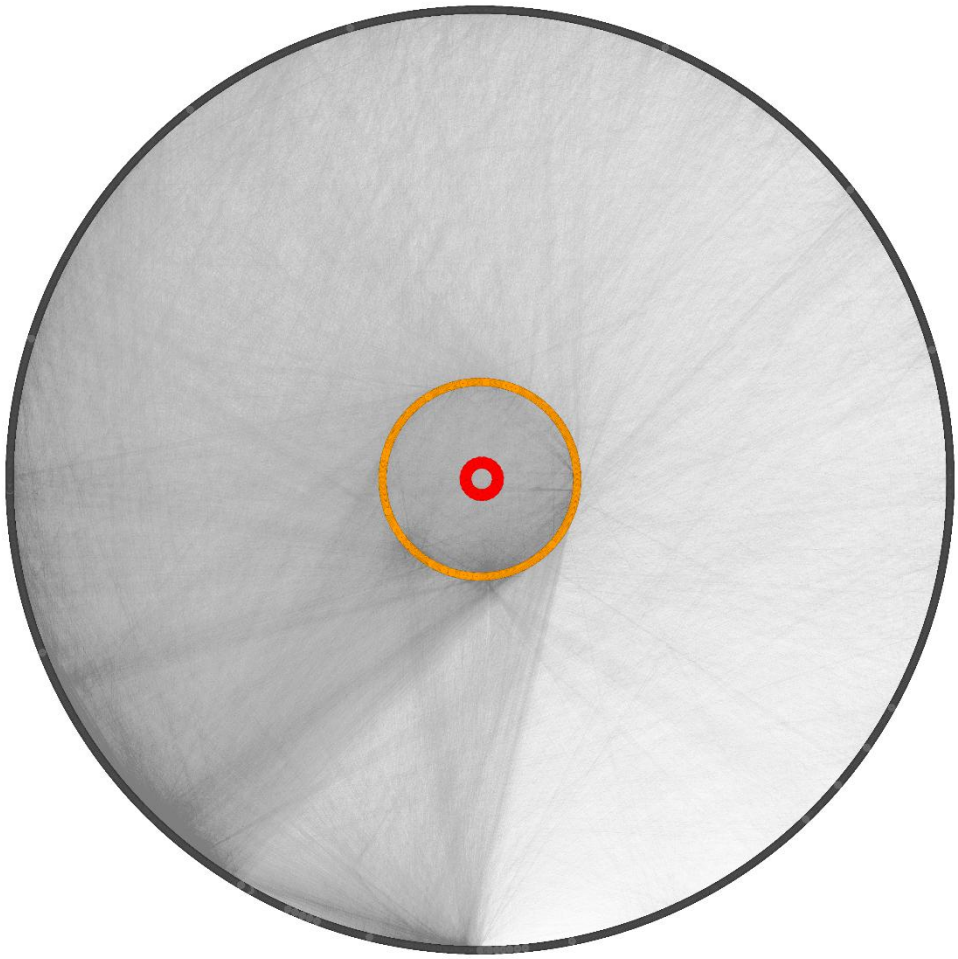


Figure 6. The whole network of ARSs/AIMPs and their neighbors. This network shows ARSs/AIMPs and first- and second-neighbor genes linked to ARSs/AIMPs either directly or indirectly. The innermost, middle, and outermost nodes represent 23 ARSs/AIMPs, 586 first-neighbor genes, and 13,539 second-neighbor genes, respectively.

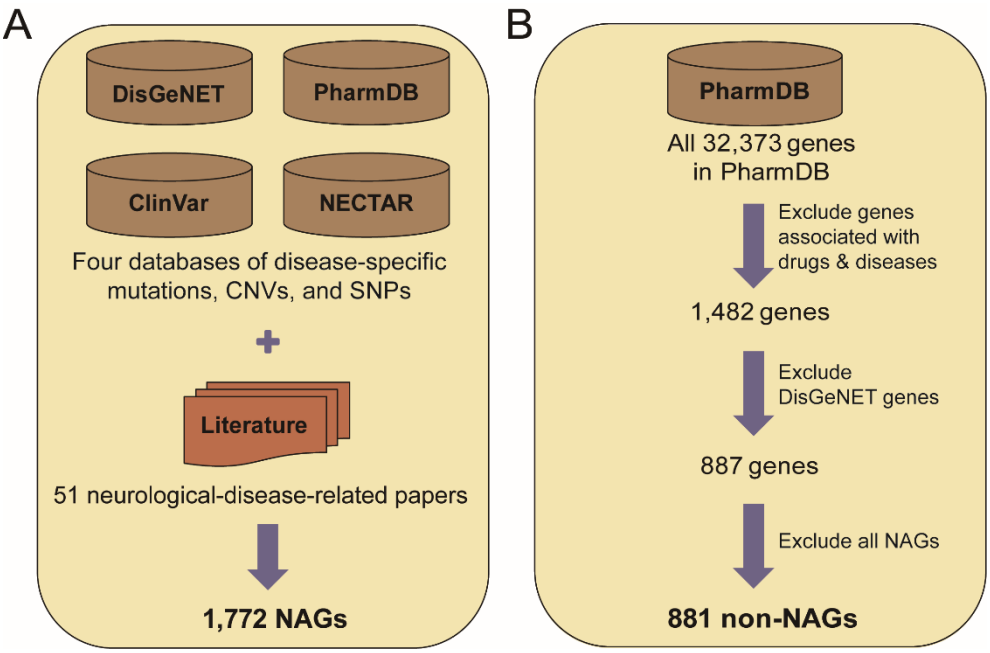
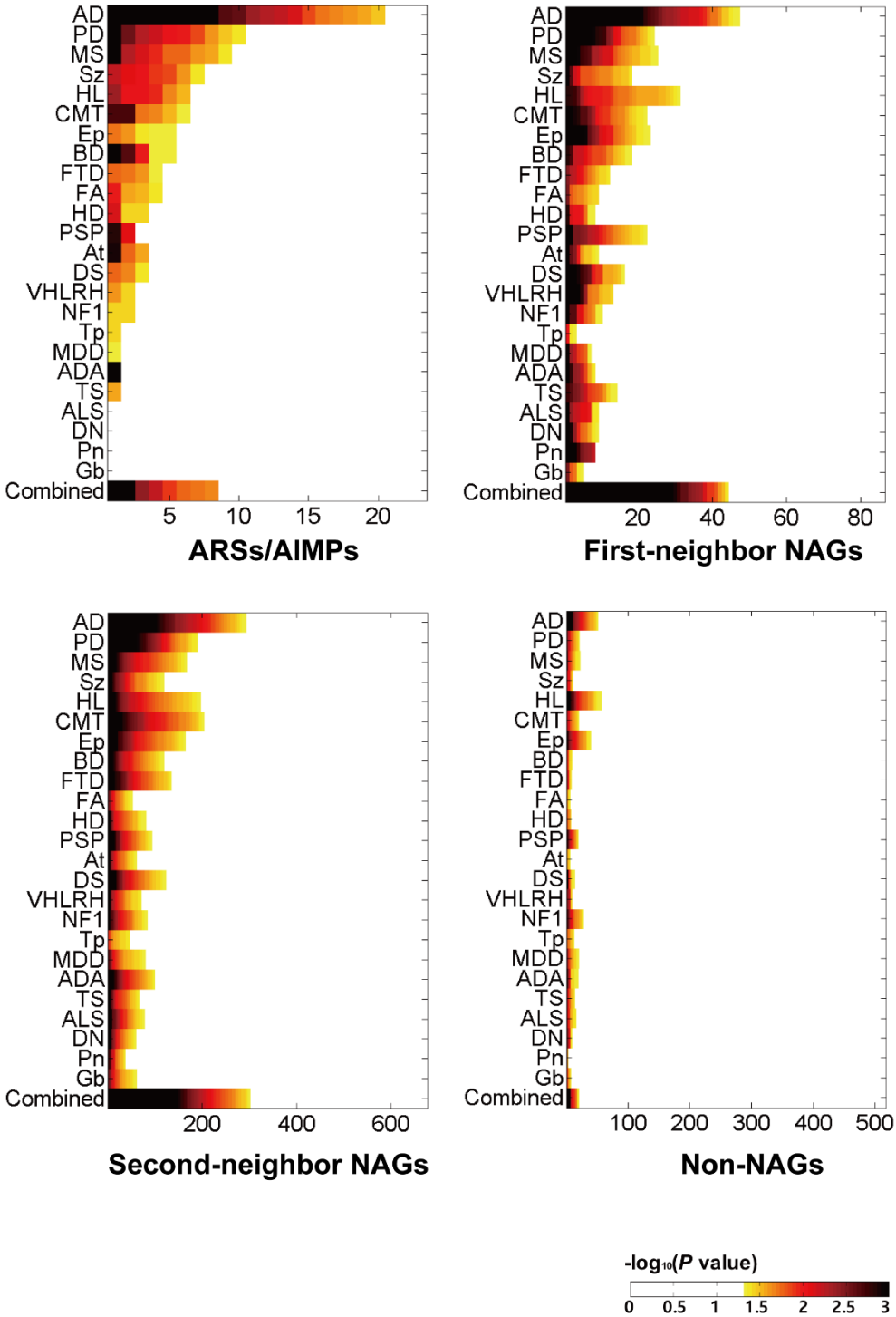


Figure 7. Selection of NAGs and non-NAGs. (A) Sources of NAGs. Four databases of disease-specific mutations, copy number variations (CNVs), and single nucleotide polymorphisms (SNPs) were used to compile the list of genes associated with 27 previously selected neurological diseases. The other NAGs were retrieved from a literature search. (B) The source of non-NAGs. A total of 881 non-NAGs were retrieved from PharmDB by means of the following criteria: (i) 1,482 genes not associated with drug actions or diseases were selected from all the 32,373 genes in PharmDB; (ii) among them, 881 non-NAGs not present either in DisGeNET or among NAGs were selected.

A


AD
PD
MS
Sz
HL
CMT
Ep
BD
FTD
FA
HD
PSP
At
DS
VHLRH
NF1
Tp
MDD
ADA
TS
ALS
DN
Pn
Gb
Combined
5
10
15
20
ARSs/AIMPs
20
40
60
80
First-neighbor NAGs
200
400
600
Second-neighbor NAGs
100
200
300
400
500
Non-NAGs
-log10(P value)
0
0.5
1
1.5
2
2.5
3

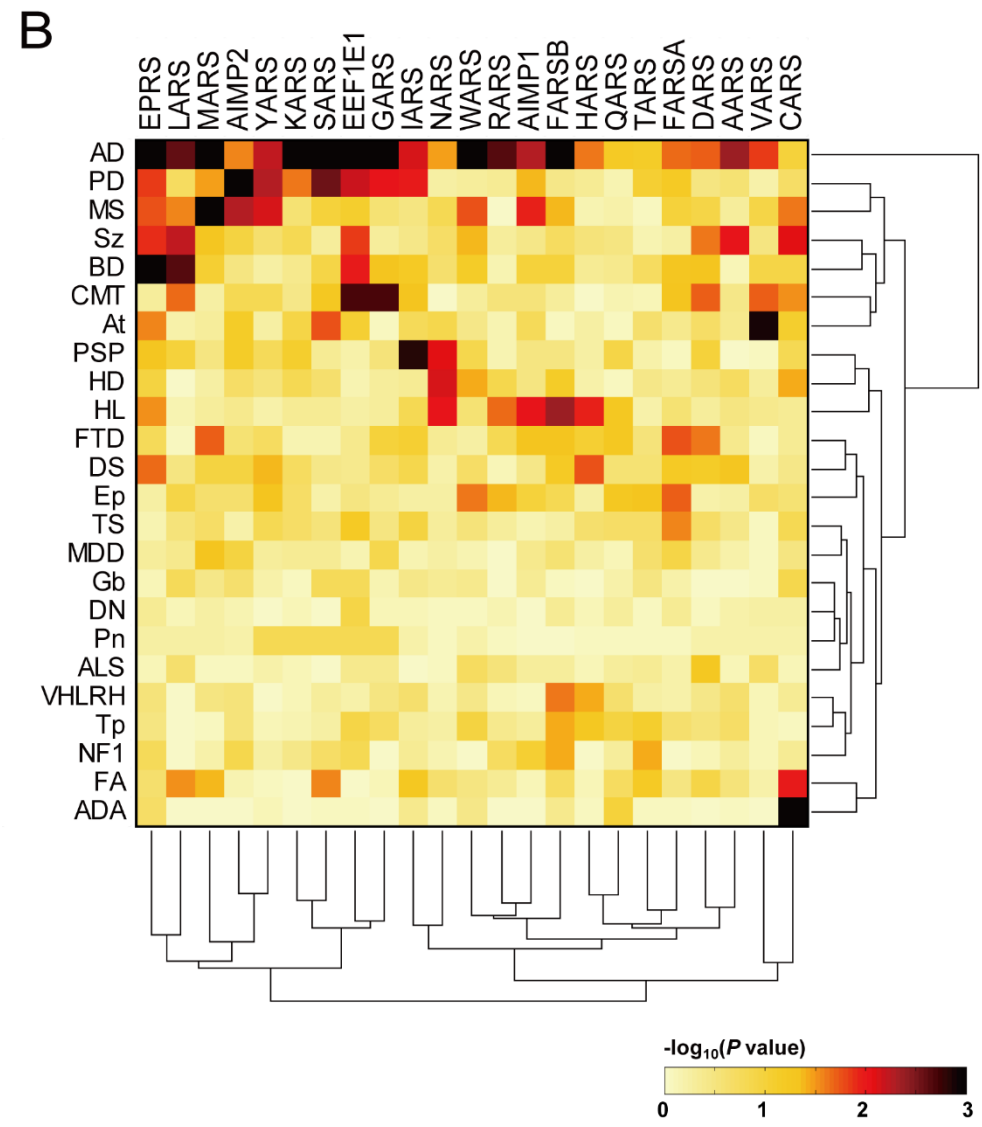


Figure 8. Proportions of DEGs in four groups and mRNA expression profiles of 23 ARSs/AIMPs in 24 neurological diseases. (A) Gene expression profiles of four groups are shown. In this figure, the x-axis shows the number of genes in each group, and the y-axis shows the diseases. Given that some of the *P* values were not available, 11 second neighbor NAGs and 364 non-NAGs were left

out. Combined *P* values of the 24 neurological diseases are shown in the bottom rows of the graphs. The white color means that $-\log_{10}$ (*P* value) is less than 1.3 and therefore the *P* value is greater than 0.05. The black color means that the *P* value is less than 0.001. (B) Expression profiles of each ARS/AIMP in the 24 neurological diseases. The heat map shows the expression profile of each ARS/AIMP in the ARS/AIMP group from Figure 8A. The 24 neurological diseases and 23 ARSs/AIMPs were grouped by *P* values (the dendrogram on the right and bottom). ARS/AIMP abbreviations comply with HUGO Gene Nomenclature Committee official symbols. For disease abbreviations, see Table 2.

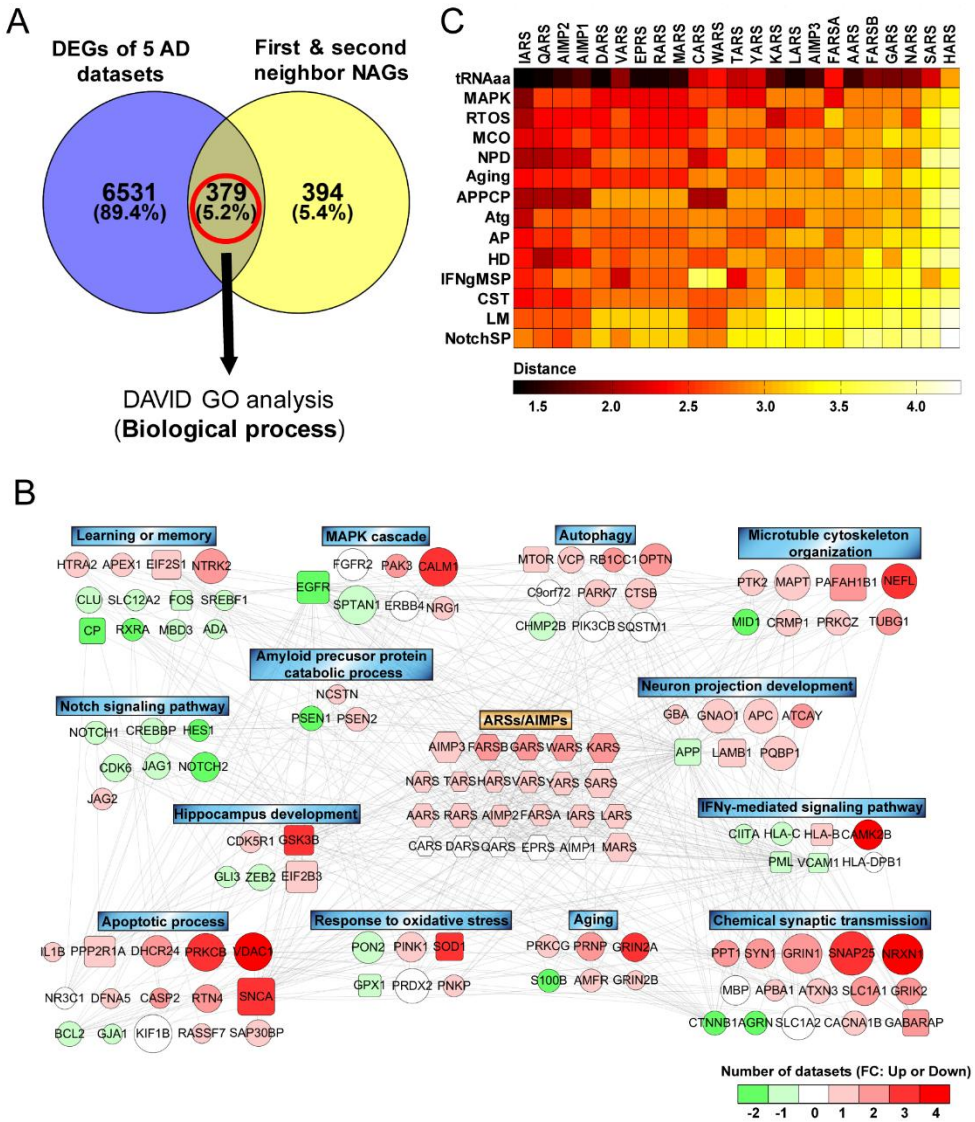


Figure 9. A hypothetical ARS/AIMP subnetwork of AD. (A) A total of 379 genes were selected as GO analysis input. All these genes are first- and second-neighbor NAGs as well as DEGs in at least one dataset. (B) This subnetwork shows that 131 genes are grouped by biological processes. GO analysis was conducted using DAVID. Hexagons, rectangles, and ellipses represent the 23 ARSs/AIMPs,

and first−, and second−neighbor NAGs, respectively. The *P* value and fold change for AD were represented by the node size and color, respectively. (C) This heat map shows the average shortest path between the 23 ARSs/AIMPs and genes included in the 13 biological processes represented in the network (B). ARS/AIMP abbreviations comply with HUGO Gene Nomenclature Committee official symbols. AP = Apoptotic process; APPCP = Amyloid precursor protein catabolic process; Atg = Autophagy; CST = Chemical synaptic transmission; HD = Hippocampus development; IFNgMSP = Interferon−γ−mediated signaling pathway; LM = Learning or memory; MAPK = MAPK cascade; MCO = Microtubule cytoskeleton organization; NotchSP = Notch signaling pathway; NPD = Neuron projection development; RTOS = Response to oxidative stress; tRNAaa = tRNA aminoacylation.

Table 1. ARS/AIMP mutations found in neurological diseases.

GeneSymbol	Disease	Position(s)	Length	Description	feature identifier	PMID
AARS	CMT2N	71	1	N → Y	VAR_067084	22206013
AARS	CMT2N	329	1	R → H	VAR_063527	20045102, 22009580
AARS	EIEE29	81	1	K → T	VAR_073719	25817015
AARS	EIEE29	751	1	R → G	VAR_073720	25817015
DARS	HBSL	256	1	M → L	VAR_070038	23643384
DARS	HBSL	274	1	A → V	VAR_070039	23643384
DARS	HBSL	367	1	D → Y	VAR_070040	23643384
DARS	HBSL	460	1	R → H	VAR_070041	23643384
DARS	HBSL	464	1	P → L	VAR_070042	23643384
DARS	HBSL	487	1	R → C	VAR_070043	23643384
DARS	HBSL	494	1	R → C	VAR_070044	23643384
DARS	HBSL	494	1	R → G	VAR_070045	23643384
GARS	CMT2D	111	1	A → V	VAR_073187	25168514
GARS	CMT2D	125	1	E → G	VAR_018718	17035524, 25168514, 12690580
GARS	CMT2D	200	1	D → N	VAR_073188	25168514
GARS	CMT2D	200	1	D → Y	VAR_074016	26244500
GARS	CMT2D	265	1	S → F	VAR_073189	25168514
GARS	CMT2D	292	1	M → R	VAR_074017	26244500
GARS	CMT2D	294	1	G → R	VAR_018720	17035524, 25168514, 12690580
GARS	CMT2D	298	1	P → L	VAR_073190	25168514
GARS	CMT2D	334	1	I → F	VAR_073191	25168514
GARS	CMT2D	554	1	D → N	VAR_073193	25168514
GARS	CMT2D	652	1	G → A	VAR_073195	25168514
GARS	HMN5A	183	1	L → P	VAR_018719	17035524, 25168514, 12690580
GARS	HMN5A	472	1	H → R	VAR_073192	17035524, 25168514, 24627108
GARS	HMN5A	580	1	G → R	VAR_018721	17035524, 25168514, 12690580, 17545306
HARS	USH3B	454	1	Y → S	VAR_067918	22279524
HARS	CMT2W	132	1	T → I	VAR_075064	26072516
HARS	CMT2W	134	1	P → H	VAR_075065	26072516
HARS	CMT2W	137	1	R → Q	VAR_069022	22930593
HARS	CMT2W	175	1	D → E	VAR_075066	26072516
HARS	CMT2W	238	1	V → A	VAR_069024	22930593
HARS	CMT2W	364	1	D → Y	VAR_075067	26072516
HARS	CMT2W	505	1	P → S	VAR_069026	22930593
KARS	CMTRIB	105	1	L → H	VAR_064911	20920668
KARS	CMTRIB	274	1	I → M	VAR_064912	20920668
KARS	DFNB89	145	1	Y → H	VAR_070233	23768514
KARS	DFNB89	349	1	D → N	VAR_070234	23768514

Continued

Table 1. Continued.

GeneSymbol	Disease	Position(s)	Length	Description	feature identifier	PMID
MARS	CMT2U	618	1	R → C	VAR_073377	23729695
MARS	CMT2U	800	1	P → T	VAR_073378	24354524
QARS	MSCCA	45	1	G → V	VAR_071189	24656866
QARS	MSCCA	57	1	Y → H	VAR_071190	24656866
QARS	MSCCA	403	1	R → W	VAR_071191	24656866
QARS	MSCCA	515	1	R → W	VAR_071192	24656866
RARS	HLD9	2	1	D → G	VAR_072666	24777941
RARS	HLD9	512	1	R → Q	VAR_072667	24777941
YARS	CMTDIC	41	1	G → R	VAR_026681	16429158
YARS	CMTDIC	153-156	4	Missing	VAR_026682	16429158
YARS	CMTDIC	196	1	E → K	VAR_026684	16429158
AIMP1	HLD3	-	-	-	-	21092922

Mutations in ARSs/AIMPs were retrieved from UniProt. Detailed information can be found by searching for the feature identifier or PMID. CMT2N = Charcot-Marie-Tooth disease 2N; EIEE29 = Epileptic encephalopathy, early infantile, 29; HBSL = Hypomyelination with brainstem and spinal cord involvement and leg spasticity; CMT2D = Charcot-Marie-Tooth disease 2D; HMN5A = Neuronopathy, distal hereditary motor, 5A; USH3B = Usher syndrome 3B; CMT2W = Charcot-Marie-Tooth disease 2W; CMTRIB = Charcot-Marie-Tooth disease, recessive, intermediate type, B; DFNB89 = Deafness, autosomal recessive, 89; CMT2U = Charcot-Marie-Tooth disease 2U; MSCCA = Microcephaly, progressive, with seizures and cerebral and cerebellar atrophy; HLD9 = Leukodystrophy, hypomyelinating, 9; CMTDIC = Charcot-Marie-Tooth disease, dominant, intermediate type, C; HLD3 = Leukodystrophy, hypomyelinating, 3.

Table 2. Disease abbreviations.

Abbreviation	Full Name	NAGs	Datasets
AD	Alzheimer's disease	O	O
ADA	Alcohol and drug addiction	X	O
ALS	Amyotrophic lateral sclerosis	O	O
At	Autism	O	O
BD	Bipolar disorder	O	O
CA	Cerebellar ataxia	O	X
CD	Cerebellar diseases	O	X
CMT	Charcot-Marie-Tooth disease	O	O
DN	Diabetic neuropathy	X	O
DS	Down syndrome	O	O
Ep	Epilepsy	O	O
FA	Friedreich's ataxia	X	O
FTD	Frontotemporal dementia	X	O
Gb	Glioblastoma	O	O
HBSL	Hypomyelination with brainstem and spinal cord involvement and leg spasticity	O	X
HD	Huntington's disease	X	O
HL	Hearing loss	O	O
HSP	Hereditary spastic paraplegia	O	X
ID	Intellectual disability	O	X
LBD	Lewy body disease	O	X
Ln	Leukoencephalopathies	O	X
Mc	Microcephaly	O	X
MDD	Major depressive disorder	X	O
MS	Multiple sclerosis	O	O
NF1	Neurofibromatosis type 1	X	O
PD	Parkinson's disease	O	O
PH	Pontocerebellar hypoplasia	O	X
Pn	Pain	X	O
PNMHH	Peripheral neuropathy, myopathy, hoarseness, and hearing loss	O	X
PS	Perrault syndrome	O	X
PSP	Progressive supranuclear palsy	X	O
SMA	Spinal muscular atrophy	O	X
SMAPH	Spinal muscular atrophy with pontocerebellar hypoplasia	O	X
SMAPME	Spinal muscular atrophy with progressive myoclonic epilepsy	O	X
Sz	Schizophrenia	O	O
Tp	Tauopathies	X	O
TS	Tuberous sclerosis	X	O
US	Usher syndromes	O	X
VHLRH	Von Hippel-Lindau-related hemangioblastoma	X	O

For brevity, the names of the 39 neurological diseases listed above are abbreviated. These abbreviations are used throughout the text. Selected diseases for NAGs or expression datasets are marked as “O” in the “NAGs” column and “Datasets” column, respectively.

Table 3. Sources of mRNA expression datasets.

Disease	Source ID	Source Database	Platform	Organism	Sample Type	N method	Number of Samples		PMID
							H	D	
AD	GSE1297	GEO	GPL96	Homo sapiens	Brain	GC-RMA	9	22	14769913
	GSE21779	GEO	GPL570	Microcebus murinus	Barin	GC-RMA	16	2	20862281
	GSE28146	GEO	GPL570	Homo sapiens	Brain	GC-RMA	8	22	21756998
	GSE4757	GEO	GPL570	Homo sapiens	Brain	GC-RMA	10	10	16242812
	GSE5281	GEO	GPL570	Homo sapiens	Brain	GC-RMA	74	87	17077275, 18332434
ADA	GSE15774	GEO	GPL6105	Mus musculus	Brain	Quantile	23	85	20459597, 24010892
	GSE20568	GEO	GPL10117	Homo sapiens	Brain	Quantile	5	15	20477762
	GSE3016	GEO	GPL88	Rattus norvegicus	Brain	GC-RMA	21	42	16076954
	GSE7762	GEO	GPL1261	Mus musculus	Brain	GC-RMA	12	24	17598886
ALS	GSE18920	GEO	GPL5188	Homo sapiens	Spinal cord	PLIER	20	24	19864493
	GSE26276	GEO	GPL6244	Homo sapiens	Skeletal muscle	PLIER	3	3	21375368
	GSE26927	GEO	GPL6104	Homo sapiens	Cervical spinal cord	Quantile	10	10	22864814, 25119539
	GSE833	GEO	GPL80	Homo sapiens	Spinal cord	GC-RMA	4	7	14645737
At	GSE26415	GEO	GPL6480	Homo sapiens	Venous blood	Quantile	42	21	21935445
	GSE6575	GEO	GPL570	Homo sapiens	Whole blood	GC-RMA	12	35	18006270
BD	GSE5389	GEO	GPL96	Homo sapiens	Brain	GC-RMA	11	10	16894394
	GSE7036	GEO	GPL570	Homo sapiens	Lymphoblastoid cell	GC-RMA	3	3	17440432
	GSE12679	GEO	GPL570	Homo sapiens	Brain	GC-RMA	6	5	19088852
CMT	GSE7423	GEO	GPL1708	Homo sapiens	Sural nerve	Quantile	4	4	Citation missing
DN	GSE11343	GEO	GPL1261	Mus musculus	Sciatic nerve	GC-RMA	4	5	18583417
	GSE27382	GEO	GPL9746	Mus musculus	Sciatic nerve	GC-RMA	7	6	21617178
	GSE34889	GEO	GPL9746	Mus musculus	Sciatic nerve	GC-RMA	15	14	Citation missing
DS	GSE1294	GEO	GPL81	Mus musculus	Brain	GC-RMA	6	6	15138197
	GSE1294	GEO	GPL82	Mus musculus	Brain	GC-RMA	6	6	15138197
	GSE1611	GEO	GPL81	Mus musculus	Cerebellum	GC-RMA	6	6	15590701
	GSE16176	GEO	GPL570	Homo sapiens	Amniotic fluid	GC-RMA	7	7	19474297
	GSE17760	GEO	GPL1261	Mus musculus	Brain	GC-RMA	3	3	20661276
	GSE5390	GEO	GPL96	Homo sapiens	Brain	GC-RMA	8	7	17950572
Ep	GSE32534	GEO	GPL570	Homo sapiens	Brain	GC-RMA	5	5	23418513
FA	GSE15848	GEO	GPL6100	Mus musculus	Heart, skeletal muscle, and liver	Quantile	11	11	19376812
	GSE31208	GEO	GPL1261	Mus musculus	Heart	GC-RMA	4	4	Citation missing
FD	GSE13162	GEO	GPL571	Homo sapiens	Brain	GC-RMA	17	39	18223198
Gb	GSE12657	GEO	GPL8300	Homo sapiens	Glioblastoma	GC-RMA	5	7	Citation missing
	GSE24100	GEO	GPL4133	Homo sapiens	Stromal cells	Quantile	3	4	21928115
HD	GSE18551	GEO	GPL1261	Mus musculus	Brain	GC-RMA	8	10	20089533
	GSE26001	GEO	GPL1261	Mus musculus	Embryonic stem cells	GC-RMA	6	6	21536587
	GSE24250	GEO	GPL96	Homo sapiens	Whole blood	GC-RMA	6	8	21969577
HL	GSE4866	GEO	GPL1261	Mus musculus	Cochlea	GC-RMA	5	5	17363114
MDD	GSE19738	GEO	GPL6848	Homo sapiens	Whole blood	Quantile	34	33	20471630
	GSE32280	GEO	GPL570	Homo sapiens	Peripheral blood	GC-RMA	8	8	Citation missing

Continued

Table 3. Continued.

Disease	Source ID	Source Database	Platform	Organism	Sample Type	N method	Number of Samples		PMID
							H	D	
MS	GSE13732	GEO	GPL570	Homo sapiens	PBMC	GC-RMA	40	73	18689680
	GSE16461	GEO	GPL570	Homo sapiens	PBMC	GC-RMA	8	8	21216829
	GSE21942	GEO	GPL570	Homo sapiens	PBMC	GC-RMA	15	14	22021740
	GSE23832	GEO	GPL6244	Homo sapiens	PBMC	PLIER	4	8	21346816
	GSE26484	GEO	GPL570	Homo sapiens	PBMC	GC-RMA	4	6	22491253
	GSE38010	GEO	GPL570	Homo sapiens	Brain	GC-RMA	2	5	22734047
Nf1	GSE29343	GEO	GPL6246	Mus musculus	Triceps muscles	PLIER	4	4	21478499
PD	E-GEOD-20141	ArrayExpress	GPL570	Homo sapiens	Brain	GC-RMA	8	10	20926834
	E-GEOD-20146	ArrayExpress	GPL570	Homo sapiens	Brain	GC-RMA	10	10	20926834
	E-GEOD-20153	ArrayExpress	GPL570	Homo sapiens	Lymphoblasts	GC-RMA	8	8	20926834
	E-GEOD-20168	ArrayExpress	GPL96	Homo sapiens	Brain	GC-RMA	15	14	15965975, 20926834
	E-GEOD-20291	ArrayExpress	GPL96	Homo sapiens	Brain	GC-RMA	20	15	15965975, 20926834
	E-GEOD-20292	ArrayExpress	GPL96	Homo sapiens	Brain	GC-RMA	18	11	15965975, 20926834
	E-GEOD-20295	ArrayExpress	GPL96	Homo sapiens	Brain	GC-RMA	53	40	15965975, 20926834
	E-GEOD-20333	ArrayExpress	GPL201	Homo sapiens	Brain	GC-RMA	6	6	Citation missing
	E-GEOD-7621	ArrayExpress	GPL570	Homo sapiens	Brain	GC-RMA	9	16	17571925
	E-MEXP-1416	ArrayExpress	GPL1352	Homo sapiens	Brain	GC-RMA	8	8	17412603
Pn	GSE38859	GEO	GPL6543	Rattus norvegicus	Spinal cord	PLIER	6	6	Citation missing
PSP	GSE6613	GEO	GPL96	Homo sapiens	Whole blood	GC-RMA	22	6	17215369
Sz	GSE21935	GEO	GPL570	Homo sapiens	Brain	GC-RMA	19	23	21538462
	GSE34058	GEO	GPL6247	Rattus norvegicus	Fetal brain	PLIER	10	9	22310921
	GSE12679	GEO	GPL570	Homo sapiens	Brain	GC-RMA	11	16	19088852
Tp	GSE13691	GEO	GPL339	Mus musculus	Brain	GC-RMA	5	6	Citation missing
	GSE13691	GEO	GPL340	Mus musculus	Brain	GC-RMA	6	6	Citation missing
TS	GSE16969	GEO	GPL570	Homo sapiens	Brain	GC-RMA	4	6	19912235
	GSE29797	GEO	GPL11180	Mus musculus	Lymphoid organs	GC-RMA	8	8	21765414
	GSE40630	GEO	GPL6246	Mus musculus	Blood and cerebellum	PLIER	16	16	24564913
	GSE9715	GEO	GPL96	Homo sapiens	Skin	GC-RMA	4	7	18292222
VHLRH	GSE11484	GEO	GPL1261	Mus musculus	Pancreatic islets	GC-RMA	3	3	19056893
	GSE20235	GEO	GPL1261	Mus musculus	Glomerulus	GC-RMA	3	3	20522651

Platforms are specified as a GEO accession number (GPLxxx). PBMC, H, and D stand for peripheral blood mononuclear cell, healthy controls, and disease group, respectively. N method means the normalization method. Papers cited for each dataset can be found by searching for PMIDs. For abbreviations, see Table 2.

Table 4. Description of all NAGs.

Disease	Number of genes in each source DB				Total (not duplicated)
	DisGeNET	PharmDB	ClinVar or NECTAR	Literature	
ID	0	159	164	65	**309**
At	10	241	0	0	**243**
CMT	14	23	45	219	**242**
Sz	176	177	0	0	**177**
AD	83	86	33	85	**174**
PD	61	67	45	99	**172**
Ep	57	66	24	102	**171**
ALS	18	28	24	141	**160**
HL	27	70	59	37	**113**
HSP	4	5	40	79	**83**
Mc	23	30	10	37	**64**
Gb	51	50	0	0	**54**
MS	40	40	0	0	**40**
BD	38	38	0	0	**38**
SMA	4	8	13	19	**25**
PS	0	0	5	18	**19**
CD	5	5	0	13	**17**
Ln	0	14	6	2	**16**
DS	15	16	0	0	**16**
US	0	8	11	12	**13**
CA	2	7	1	2	**10**
PH	0	0	0	5	**5**
LBD	0	0	0	1	**1**
HBSL	0	0	0	1	**1**
SMAPH	0	0	0	1	**1**
SMAPME	0	0	0	1	**1**
PNMHH	0	0	0	1	**1**
Total (not duplicated)	**533**	**934**	**453**	**866**	1772

We collected 1,772 genes associated with 27 neurological diseases from DisGeNET, PharmDB, ClinVar, NECTAR, and the literature. Total numbers of genes associated with each disease were sorted in descending order. For abbreviations, see Table 2.

Discussion

Even though many studies on neurological diseases associated with ARSs/AIMPs have been published,[3,85] there are still many other neurological diseases ARSs/AIMPs can be implicated in. Therefore, we included neurological diseases which are not well-known for a connection with ARSs/AIMPs. Charcot-Marie-Tooth (CMT) diseases can be considered a neurodegenerative disease affecting the peripheral nervous system. Furthermore, many mutations of ARSs (alanyl-, glycyl-, histidyl-, lysyl-, methionyl- and tyrosyl-tRNA synthetase) have been found in this disease.[14,86] Therefore, we included other neurodegenerative diseases affecting the central nervous system, e.g. AD, Parkinson's disease, and Huntington's disease. Differentially expressed ARSs/AIMPs and their neighbors were the most numerous in AD. Consequently, we needed to determine how ARSs/AIMPs are involved in the pathophysiology or pathogenesis of AD.

AD is a representative neurodegenerative disease and there is no cure yet.[87] Extracellular amyloid-β (Aβ) plaques and neurofibrillary tangles composed of hyperphosphorylated tau protein are found in the brain of AD patients. Amyloid precursor protein (*APP*) is a membrane protein and can be cleaved by α-secretase. Nevertheless, *APP* can also be cleaved by β-secretase. If *APP* is cleaved by β-secretase, the remaining membrane-

anchored fragment known as C99 can be cleaved by γ-secretase thus releasing insoluble Aβ.[88] In an experiment on protein arrays, more than 2,200 proteins were found to bind to Aβ including six ARSs/AIMPs (isoleucyl-, glutaminyl-, tyrosyl-, cysteinyl-tRNA synthetase, *AIMP1*, and *AIMP2*).[89] Because Aβ is also found inside neurons,[88] ARSs/AIMPs can bind to intracellular Aβ. Consequently, ARSs/AIMPs bound to Aβs may show some alterations in their canonical function or an unknown non-canonical function in neurons. In addition, ARSs/AIMPs can be secreted from the cell[5,90]; therefore, these proteins can be involved in the formation of extracellular Aβ plaques.

Recently, a network-based approach to disease research has been receiving much attention.[91] As we mentioned earlier, however, except for cancer, there has been little biological-network-based research on the pathological role of ARSs. Therefore, we compiled a subnetwork showing AD-associated biological processes using DAVID. Among 13 biological processes, "MAPK cascade" and "response to oxidative stress" were the nearest to ARSs/AIMPs when we calculated average shortest paths. It is reported that reactive oxygen species facilitate activation of MAPK signaling and hyperphosphorylation of tau protein.[92,93] In our study, glutathione peroxidase 1 (*GPX1*) was found to be downregulated, and this protein protects cells from oxidative stress by reducing hydrogen peroxide to water. Furthermore, according to the subnetwork that we created, *GPX1* is directly connected to

KARS. Therefore, we need to confirm experimentally that *KARS* is indirectly involved in oxidative stress through *GPX1*.

Glycogen synthase kinase 3 (GSK3) is a serine/threonine kinase and there are two isoforms of this protein (GSK3α and GSK3β). Many studies have shown that GSK3 is associated with AD.[94,95] Eventually, those researcher advanced the "GSK3 hypothesis" and according to this theory, tau protein phosphorylation, increased Aβ production, memory loss, and inflammation are associated with GSK3.[94,95] Three up-regulated ARSs or AIMPs (*AIMP2*, *LARS*, *RARS*) are indirectly connected to GSK3β through *RPS6KA1* and GSK3 is inhibited via phosphorylation by *RPS6KA1*[96,97]. Therefore, ARSs/AIMPs can be implicated in the GSK3 hypothesis because it is possible that they regulate the activity of *RPS6KA1*.

In conclusion, we confirmed that the 23 ARSs/AIMPs and their neighbors are differentially expressed, especially in AD. Therefore, further research on the role of ARSs/AIMPs in AD is needed.

References

1. Park, S.G., et al., Aminoacyl-tRNA synthetase-interacting multifunctional proteins (AIMPs): a triad for cellular homeostasis. *IUBMB Life*, 2010. 62(4): p. 296-302.
2. Han, J.M., et al., Hierarchical network between the components of the multi-tRNA synthetase complex: Implications for complex formation. *Journal of Biological Chemistry*, 2006. 281(50): p. 38663-38667.
3. Park, S.G., et al., Aminoacyl tRNA synthetases and their connections to disease. *Proceedings of the National Academy of Sciences of the United States of America*, 2008. 105(32): p. 11043-11049.
4. Park, S.G., et al., Functional expansion of aminoacyl-tRNA synthetases and their interacting factors: new perspectives on housekeepers. *Trends in Biochemical Sciences*, 2005. 30(10): p. 569-74.
5. Kim, S., et al., Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. *Nature Reviews Cancer*, 2011. 11(10): p. 708-718.
6. Lee, J.H., et al., Comprehensive data resources and analytical tools for pathological association of aminoacyl tRNA synthetases with cancer. *Database-the Journal of Biological Databases and Curation*, 2015.
7. Yao, P., et al., Aminoacyl-tRNA synthetases in medicine and disease. *EMBO Molecular Medicine*, 2013. 5(3): p. 332-43.

8. Lee, Y., et al., Parthanatos mediates AIMP2-activated age-dependent dopaminergic neuronal loss. *Nature Neuroscience*, 2013. 16(10): p. 1392-400.
9. Renton, A.E., et al., State of play in amyotrophic lateral sclerosis genetics. *Nature Neuroscience*, 2014. 17(1): p. 17-23.
10. Trushina, E., et al., Identification of Altered Metabolic Pathways in Plasma and CSF in Mild Cognitive Impairment and Alzheimer's Disease Using Metabolomics. *PloS One*, 2013. 8(6).
11. Lee, J.Y., et al., Promiscuous methionyl-tRNA synthetase mediates adaptive mistranslation to protect cells against oxidative stress. *Journal of Cell Science*, 2014. 127(Pt 19): p. 4234-45.
12. Ling, J., et al., Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site. *Proceedings of the National Academy of Sciences of the United States of America*, 2010. 107(9): p. 4028-33.
13. Nussbaum, J.M., et al., Alzheimer disease: a tale of two prions. *Prion*, 2013. 7(1): p. 14-9.
14. Abbott, J.A., et al., Transfer RNA and human disease. *Frontiers in genetics*, 2014. 5: p. 158.
15. Lee, H.S., et al., Rational drug repositioning guided by an integrated pharmacological network of protein, disease and drug. *BMC Systems Biology*, 2012. 6: p. 80.
16. Pinero, J., et al., DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes.

Database: The Journal of Biological Databases and Curation, 2015. 2015: p. bav028.

17. Landrum, M.J., et al., ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, 2014. 42(D1): p. D980-D985.
18. Gong, S., et al., NECTAR: a database of codon-centric missense variant annotations. *Nucleic Acids Research*, 2014. 42(D1): p. D1013-D1019.
19. Myers, C.T., et al., Advancing epilepsy genetics in the genomic era. *Genome Medicine*, 2015. 7: p. 91.
20. Prihar, G., et al., Alzheimer disease PS-1 exon 9 deletion defined. *Nature Medicine*, 1999. 5(10): p. 1090.
21. Swaminathan, S., et al., Analysis of copy number variation in Alzheimer's disease: the NIALOAD/ NCRAD Family Study. *Current Alzheimer Research*, 2012. 9(7): p. 801-14.
22. Woods, C.G., et al., Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. *American Journal of Human Genetics*, 2005. 76(5): p. 717-28.
23. Mahmood, S., et al., Autosomal Recessive Primary Microcephaly (MCPH): clinical manifestations, genetic heterogeneity and mutation continuum. *Orphanet Journal of Rare Diseases*, 2011. 6: p. 39.
24. Mc Cormack, A., et al., A case of 17q21.31 microduplication and 7q31.33 microdeletion, associated with developmental delay, microcephaly, and mild dysmorphic features. *Case reports in genetics*, 2014. 2014: p. 658570.
25. Namavar, Y., et al., Classification, diagnosis and potential

mechanisms in pontocerebellar hypoplasia. *Orphanet Journal of Rare Diseases*, 2011. 6: p. 50.

26. Darvish, H., et al., A clinical and molecular genetic study of 112 Iranian families with primary microcephaly. *Journal of Medical Genetics*, 2010. 47(12): p. 823-8.

27. Hu, J., et al., CNTN6 copy number variations in 14 patients: a possible candidate gene for neurodevelopmental and neuropsychiatric disorders. *Journal of Neurodevelopmental Disorders*, 2015. 7(1): p. 26.

28. Shearer, A.E., et al., Copy number variants are a common cause of non-syndromic hearing loss. *Genome Medicine*, 2014. 6(5): p. 37.

29. Olson, H., et al., Copy number variation plays an important role in clinical epilepsy. *Annals of Neurology*, 2014. 75(6): p. 943-58.

30. Amara, A., et al., Correlation of SMN2, NAIP, p44, H4F5 and Occludin genes copy number with spinal muscular atrophy phenotype in Tunisian patients. *European Journal of Paediatric Neurology*, 2012. 16(2): p. 167-74.

31. Kaufman, L., et al., The genetic basis of non-syndromic intellectual disability: a review. *Journal of Neurodevelopmental Disorders*, 2010. 2(4): p. 182-209.

32. Puffenberger, E.G., et al., Genetic mapping and exome sequencing identify variants associated with five novel diseases. *PloS One*, 2012. 7(1): p. e28936.

33. Steinlein, O.K., Genetic mechanisms that underlie epilepsy. *Nature Reviews Neuroscience*, 2004. 5(5): p. 400-408.

34. Shafique, S., et al., Genetic spectrum of autosomal recessive

non-syndromic hearing loss in Pakistani families. *PloS One*, 2014. 9(6): p. e100146.

35. Ramos, E.I., et al., Genetic variation in MKL2 and decreased downstream PCTAIRE1 expression in extreme, fatal primary human microcephaly. *Clinical Genetics*, 2014. 85(5): p. 423-32.

36. Bekris, L.M., et al., Genetics of Alzheimer disease. *Journal of Geriatric Psychiatry and Neurology*, 2010. 23(4): p. 213-27.

37. Bertram, L., et al., The genetics of Alzheimer's disease. *Progress in Molecular Biology and Translational Science*, 2012. 107: p. 79-100.

38. Chen, S., et al., Genetics of amyotrophic lateral sclerosis: an update. *Molecular Neurodegeneration*, 2013. 8: p. 28.

39. Klein, C., et al., Genetics of Parkinson's disease. *Cold Spring Harbor Perspectives in Medicine*, 2012. 2(1): p. a008888.

40. Wee, C.D., et al., The genetics of spinal muscular atrophies. *Current Opinion in Neurology*, 2010. 23(5): p. 450-8.

41. Schymick, J.C., et al., Genetics of sporadic amyotrophic lateral sclerosis. *Human Molecular Genetics*, 2007. 16 Spec No. 2: p. R233-42.

42. Uyan, O., et al., Genome-wide copy number variation in sporadic amyotrophic lateral sclerosis in the Turkish population: deletion of EPHA3 is a possible protective factor. *PloS One*, 2013. 8(8): p. e72381.

43. Lee, J.A., et al., Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*, 2006. 52(1): p. 103-121.

44. Kim, M.J., et al., Genotype and phenotype heterogeneity in perrault syndrome. *Journal of Pediatric and Adolescent*

Gynecology, 2013. 26(1): p. e25-7.

45. Salinas, S., et al., Hereditary spastic paraplegia: clinical features and pathogenetic mechanisms. *Lancet Neurology*, 2008. 7(12): p. 1127-38.

46. Lo Giudice, T., et al., Hereditary spastic paraplegia: clinical-genetic characteristics and evolving molecular mechanisms. *Experimental Neurology*, 2014. 261: p. 518-39.

47. Thomas, R.H., et al., The hidden genetics of epilepsy-a clinically important new paradigm. *Nature Reviews: Neurology*, 2014. 10(5): p. 283-92.

48. Liu, X., et al., Increased Rate of Sporadic and Recurrent Rare Genic Copy Number Variants in Parkinson's Disease Among Ashkenazi Jews. *Molecular genetics & genomic medicine*, 2013. 1(3): p. 142-154.

49. Scheper, G.C., et al., Mitochondrial aspartyl-tRNA synthetase deficiency causes leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation. *Nature Genetics*, 2007. 39(4): p. 534-9.

50. Irobi, J., et al., Molecular genetics of distal hereditary motor neuropathies. *Human Molecular Genetics*, 2004. 13 Spec No 2: p. R195-202.

51. Reddy, R., et al., Molecular genetics of the Usher syndrome in Lebanon: identification of 11 novel protein truncating mutations by whole exome sequencing. *PloS One*, 2014. 9(9): p. e107326.

52. Ahmed, Z.M., et al., The molecular genetics of Usher syndrome. *Clinical Genetics*, 2003. 63(6): p. 431-44.

53. Fogel, B.L., et al., Mutations in rare ataxia genes are

uncommon causes of sporadic cerebellar ataxia. *Movement Disorders*, 2012. 27(3): p. 442-6.

54. Lieber, D.S., et al., Next generation sequencing with copy number variant detection expands the phenotypic spectrum of HSD17B4-deficiency. *BMC Medical Genetics*, 2014. 15: p. 30.

55. Lesage, S., et al., Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Human Molecular Genetics*, 2009. 18(R1): p. R48-59.

56. Jenkinson, E.M., et al., Perrault Syndrome Is Caused by Recessive Mutations in CLPP, Encoding a Mitochondrial ATP-Dependent Chambered Protease. *American Journal of Human Genetics*, 2013. 92(4): p. 605-613.

57. Rudnik-Schoneborn, S., et al., Pontocerebellar hypoplasia. *American Journal of Medical Genetics. Part C: Seminars in Medical Genetics*, 2014. 166C(2): p. 173-83.

58. Maricich, S.M., et al., Pontocerebellar Hypoplasia: Review of Classification and Genetics, and Exclusion of Several Genes Known to Be Important for Cerebellar Development. *Journal of Child Neurology*, 2011. 26(3): p. 288-294.

59. Hooli, B.V., et al., Rare autosomal copy number variations in early-onset familial Alzheimer's disease. *Molecular Psychiatry*, 2014. 19(6): p. 676-681.

60. Beetz, C., et al., REEP1 mutation spectrum and genotype/phenotype correlation in hereditary spastic paraplegia type 31. *Brain*, 2008. 131: p. 1078-1086.

61. Wain, L.V., et al., The Role of Copy Number Variation in Susceptibility to Amyotrophic Lateral Sclerosis: Genome-Wide Association Study and Comparison with Published Loci. *PloS*

One, 2009. 4(12).

62. Russman, B.S., Spinal muscular atrophy: Clinical classification and disease heterogeneity. *Journal of Child Neurology*, 2007. 22(8): p. 946-951.

63. Ogino, S., et al., Spinal muscular atrophy: molecular genetics and diagnostics. *Expert Review of Molecular Diagnostics*, 2004. 4(1): p. 15-29.

64. Aparisi, M.J., et al., Targeted next generation sequencing for molecular diagnosis of Usher syndrome. *Orphanet Journal of Rare Diseases*, 2014. 9.

65. Millan, J.M., et al., An Update on the Genetics of Usher Syndrome. *Journal of ophthalmology*, 2011.

66. Kremer, H., et al., Usher syndrome: molecular links of pathogenesis, proteins and pathways. *Human Molecular Genetics*, 2006. 15: p. R262-R270.

67. Keats, B.J., et al., The usher syndromes. *American Journal of Medical Genetics*, 1999. 89(3): p. 158-66.

68. Lupski, J.R., et al., Whole-Genome Sequencing in a Patient with Charcot-Marie-Tooth Neuropathy. *New England Journal of Medicine*, 2010. 362(13): p. 1181-1191.

69. Chatr-Aryamontri, A., et al., The BioGRID interaction database: 2015 update. *Nucleic Acids Research*, 2015. 43(Database issue): p. D470-8.

70. Kerrien, S., et al., The IntAct molecular interaction database in 2012. *Nucleic Acids Research*, 2012. 40(Database issue): p. D841-6.

71. Barrett, T., et al., NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Research*, 2013.

41(Database issue): p. D991-5.

72. Kolesnikov, N., et al., ArrayExpress update--simplifying data submissions. *Nucleic Acids Research*, 2015. 43(Database issue): p. D1113-6.

73. Wu, Z.J., et al., A model-based background adjustment for oligonucleotide expression arrays. *Journal of the American Statistical Association*, 2004. 99(468): p. 909-917.

74. Therneau, T.M., et al., What does PLIER really do? *Cancer Informatics*, 2008. 6: p. 423-31.

75. Bolstad, B.M., et al., A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 2003. 19(2): p. 185-93.

76. Hwang, D., et al., A data integration methodology for systems biology: experimental verification. *Proceedings of the National Academy of Sciences of the United States of America*, 2005. 102(48): p. 17302-7.

77. Eppig, J.T., et al., The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Research*, 2012. 40(Database issue): p. D881-6.

78. Shannon, P., et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 2003. 13(11): p. 2498-504.

79. Burnstock, G., Purinergic signalling and disorders of the central nervous system. *Nature Reviews: Drug Discovery*, 2008. 7(7): p. 575-90.

80. Ahmed, S.S., et al., Systems biological approach on neurological disorders: a novel molecular connectivity to aging

and psychiatric diseases. *BMC Systems Biology*, 2011. 5: p. 6.

81. Altar, C.A., et al., Target identification for CNS diseases by transcriptional profiling. *Neuropsychopharmacology*, 2009. 34(1): p. 18-54.

82. Chico, L.K., et al., Targeting protein kinases in central nervous system disorders. *Nature Reviews: Drug Discovery*, 2009. 8(11): p. 892-909.

83. Foo, J.N., et al., Whole-genome and whole-exome sequencing in neurological diseases. *Nature Reviews: Neurology*, 2012. 8(9): p. 508-17.

84. Barabasi, A.L., et al., Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 2004. 5(2): p. 101-U15.

85. Antonellis, A., et al., The role of aminoacyl-tRNA synthetases in genetic diseases. *Annual Review of Genomics & Human Genetics*, 2008. 9: p. 87-107.

86. Gonzalez, M., et al., Exome sequencing identifies a significant variant in methionyl-tRNA synthetase (MARS) in a family with late-onset CMT2. *Journal of Neurology, Neurosurgery and Psychiatry*, 2013. 84(11): p. 1247-9.

87. Huang, Y., et al., Alzheimer mechanisms and therapeutic strategies. *Cell*, 2012. 148(6): p. 1204-22.

88. LaFerla, F.M., et al., Intracellular amyloid-beta in Alzheimer's disease. *Nature Reviews: Neuroscience*, 2007. 8(7): p. 499-509.

89. Olah, J., et al., Interactions of pathological hallmark proteins: tubulin polymerization promoting protein/p25, beta-amyloid, and alpha-synuclein. *Journal of Biological Chemistry*, 2011.

286(39): p. 34088-100.

90. Mirando, A.C., et al., Regulation of angiogenesis by aminoacyl-tRNA synthetases. *International Journal of Molecular Sciences*, 2014. 15(12): p. 23725-48.

91. Barabasi, A.L., et al., Network medicine: a network-based approach to human disease. *Nature Reviews: Genetics*, 2011. 12(1): p. 56-68.

92. Formichi, P., et al., Cerebrospinal fluid tau, A beta, and phosphorylated tau protein for the diagnosis of Alzheimer's disease. *Journal of Cellular Physiology*, 2006. 208(1): p. 39-46.

93. Kim, E.K., et al., Pathological roles of MAPK signaling pathways in human diseases. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2010. 1802(4): p. 396-405.

94. Hooper, C., et al., The GSK3 hypothesis of Alzheimer's disease. *Journal of Neurochemistry*, 2008. 104(6): p. 1433-9.

95. Llorens-Martin, M., et al., GSK-3beta, a pivotal kinase in Alzheimer disease. *Frontiers in Molecular Neuroscience*, 2014. 7: p. 46.

96. Sutherland, C., et al., The alpha-isoform of glycogen synthase kinase-3 from rabbit skeletal muscle is inactivated by p70 S6 kinase or MAP kinase-activated protein kinase-1 in vitro. *FEBS Letters*, 1994. 338(1): p. 37-42.

97. Cohen, P., et al., The renaissance of GSK3. *Nature Reviews: Molecular Cell Biology*, 2001. 2(10): p. 769-76.

국문초록

Aminoacyl-tRNA synthetase (ARS)는 20개의 아미노산 각각을 그에 상응하는 tRNA에 결합시켜주는 효소이다. 그러나 이들은 또한 많은 신호전달에 관여한다는 것이 알려져 있다. 이런 다양한 역할 때문에 이들은 다양한 질병에 병리학적으로 관련되어 있다. 또한 많은 ARSs 돌연변이들이 다양한 신경질환에서 발견되었다. 우리는 체계적으로 ARSs와 이들과 관련되어 있는 ARS-interacting multi-functional proteins (AIMPs)가 다양한 신경질환과 병리학적인 연관성이 있는지를 조사하였다. 3개의 protein-protein interaction (PPI) 데이터베이스를 이용하여 ARSs/AIMPs와 이들과 상호작용하는 단백질들의 전체 네트워크를 만들었다. 이 네트워크에서 586개는 ARSs/AIMPs와 직접적으로 연결되어 있는 first-neighbor이고, 13,539개는 first-neighbor를 통해 간접적으로 ARSs/AIMPs에 연결되어 있는 second-neighbor이다. 또한 공개 데이터베이스와 문헌조사를 통해 27개의 신경질환과 관련된 1,772개의 유전자 (NAGs)를 선택하였고, 이들 중에 first-neighbor에 해당하는 유전자들을 first-neighbor NAGs라고 하였다. 마찬가지로 second-neighbor 중에 NAGs에 해당하며 first-neighbor NAGs에 연결된 유전자들을 second-neighbor NAGs라고 하였다.

Gene Expression Omnibus (GEO)와 ArrayExpress에서 24개의 신경 질환과 관련된 67개의 유전자 발현 데이터세트를 모았다. ARSs/AIMPs와 그들과 PPI로 연결된 NAGs의 발현 프로파일을 음성대조군인 non-NAGs와 비교하였다. 각 데이터세트의 유전자 별 P value를 계산한 뒤, 같은 질병끼리 P value를 합쳤다. 각 질병 별 P value를

다시 하나의 P value로 합쳐서 24개의 신경 질환에서 전반적으로 어떤 유전자가 유의미하게 달리 발현되는지 알아보았다. 그리고 Database for Annotation, Visualization, and Integrated Discovery (DAVID)를 이용하여 ARSs와 그들과 연결된 유전자들이 어떤 생물학적 과정에 관여하는지, 그리고 건강한 그룹과 병에 걸린 그룹에서 유전자 발현의 차이를 P value로 보여주는 서브네트워크를 만들었다.

알츠하이머병이나 파킨슨병과 같은 신경퇴행성질환에서 상당수의 ARSs/AIMPs와 first-, second-neighbor NAGs가 건강한 그룹과 병에 걸린 그룹에서 유의미한 발현 차이를 보이는 유전자 (DEGs)였다. 요약하자면 20개의 세포질 ARSs와 3개의 AIMPs는 다양한 NAGs와 집중적으로 연결되어 있으며 신경질환에서 차등적인 발현을 보였기 때문에 신경질환과 병리학적으로 밀접한 관계가 있을 것으로 예상된다.

주요어: Aminoacyl-tRNA synthetase (ARS), 신경질환, 메신저 RNA 발현 프로파일, DEGs, Protein-protein interaction (PPI), 네트워크, GOBP (Gene ontology biological process), NAGs

학번: 2014-24873