Statistical omics data integration identifies potential therapeutic targets for synucleinopathies

Said el Bouhaddani[a,*], Matthias Höllerhage[b,*], Hae-Won Uh[a], Marc Bickle[c], Claudia Moebius[d], Günter U. Höglinger[b,e,f,**], Jeanine J. Houwing-Duistermaat[a,g,**]

[a] *Dept. of Data Science and Biostatistics, Division Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Heidelberglaan 100, 3584CX, Utrecht, Netherlands*
[b] *Department of Neurology, Hannover Medical School, Carl-Neuberg-Str. 1, 30625, Hannover, Germany*
[c] *Roche Institute for Translational Bioengineering , Grenzacherstrasse 124, 4070 Basel, Switzerland, Germany*
[d] *Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307, Dresden, Germany*
[e] *Department of Neurology, Ludwig-Maximilians-Universität, Marchioninistr. 15, 81377, Munich, Germany*
[f] *German Center for Neurodegenerative Diseases, Feodor-Lynen-Str. 17, 81377, Munich, Germany*
[g] *Dept. of Mathematics, Radboud University, Heyendaalseweg 135, 6525AJ, Nijmegen, Netherlands*

Abstract

Synucleinopathies, such as Parkinson's disease and multiple system atrophy, are characterized by α-synuclein aggregation in the brain. There is no disease-modifying therapy against synucleinopathies yet. To understand the neurobiological events underlying α-synuclein aggregation, LUHMES cell lines have been utilized to measure transcriptomics and proteomics data and screen 1600 FDA-approved drugs for their toxicity-reducing ability. However, rather than an integrative approach, the statistical analyses were performed per dataset.

We propose a novel statistical data integration method, POPLS-DA, to jointly model the transcriptomics and proteomics data and discriminate disease cells from controls. We also performed a dose-response test to validate screened FDA-approved drugs. We integrated the validated drugs with the selected genes/proteins using integrative bioinformatic analyses.

We identified clusters of synaptic and lysosome-related genes targeted by the validated drugs. Furthermore, *HSPA5*, a member of the heat shock protein 70 family, was one of the most targeted genes. HSPA5 has been previously linked to α-synuclein pathology, showing the relevance of our findings.

Our integrative approach results provide new directions for therapeutic targets for synucleinopathies.

Keywords: Parkinson's disease, Multiple system atrophy, Heat shock proteins,

[*] Shared first author. Corresponding author: Said el Bouhaddani `s.elbouhaddani@umcutrecht.nl`
[**] Shared last author.

α-synuclein, Cell lines, Latent variable modeling, Probabilistic data integration, Interactome

1. Introduction

Synucleinopathies comprise neurodegenerative disorders, the common denominator of which is the pathological aggregation of α-synuclein, encoded by the *SNCA* gene. The most common synucleinopathy is Parkinson's disease (PD), characterized by the presence of α-synuclein aggregates in the form of Lewy bodies in neuronal cells. Clinically, PD is characterized by the presence of bradykinesia in combination with rigidity and/or tremor at rest. Another neurodegenerative synucleinopathy is multiple system atrophy (MSA) that affects the nigrostriatal system, the cerebellum, and the autonomic nervous system, with a typical survival time of 6 to 10 years. Clinically, MSA is characterized by autonomic dysfunction, as well as parkinsonism or cerebellar symptoms, which determine the parkinsonian (MSA-P) and the cerebellar subtype (MSA-C). In MSA, proteinaceous inclusions of α-synuclein can be found in oligodendrocytes, termed glial cytoplasmic inclusions, and in neurons. In general, aggregated forms of α-synuclein are assumed to be toxic, leading to cell death and neurodegeneration. The mechanisms of how α-synuclein confers toxicity are not yet fully understood. Current therapies for synucleinopathies are purely symptomatic, and anti-Parkinsonian drugs are much less effective for MSA than for PD. A neuroprotective or disease modifying therapy for synucleinopathies is yet to be found. Integrative multi-omics approaches can provide insights into the biological processes, molecular functions and interactions underlying the diseases. Their ultimate goal is to identify biomarkers and promising drug targets.

LUHMES cells, which are derived from human embryonic dopaminergic cells [24], are a rich source to study the neurobiological processes underlying neurodegeneration. A recent study used LUHMES cells overexpressing α-synuclein or GFP as a control for omics research [15]. Transcriptomic and proteomic differential expression analyses have been performed. In both analyses, statistically significant results were found, but little overlap between the results were found. Enrichment analysis of the transcriptomic data identified genes involved in cell death and apoptosis, showing regulation of cell death-associated genes in the affected cells as a consequence of α-synuclein overexpression. Enrichment analyses of the proteomics data identified gene ontology terms "vesicular trafficking", "synapses" and "lysosomes". These terms have been linked to PD. Since the analyses were performed on each omics dataset separately, a holistic overview of gene deregulation in α-synuclein overexpressing cells is lacking. Such an overview should be obtained with a multi-omics approach [35], to provide a unified list of the most relevant features. Moreover, this list can comprise features that were not found with single omics analyses, because of a lack of statistical evidence when analyzed separately.

To identify drugs that protect against α-synuclein-induced cell death, LUHMES cells were used to screen 1600 FDA-approved drugs. The primary screening was performed in triplicates in two different concentrations (3 μM and 10 μM) and revealed 53 drugs that reduced cell death in LUHMES cells overexpressing α-synuclein in at least one of

the screening runs [14]. In the present study, we performed a dose-response testing of all 53 compounds to validate the protective efficacy of the drug compounds. We proposed a novel multi-omics data analysis, where we estimated the most relevant genes and proteins to distinguish between α-synuclein overexpressed and control cells. We integrated the validation results and the multi-omics integration results in an interactome network.

For the analysis of multiple omics data from cell lines, the multivariate integrative (MINT) method is a candidate method. MINT is an extension of partial least squares discriminant analysis (PLS-DA) [38] for multiple omics levels. It estimates linear combinations of features that maximally discriminate between experimental groups across all omics datasets. However, this method does not model omics-specific variation. For heterogeneous datasets, several studies have shown that modeling omics-specific variation yields better results [42, 6]. To integrate transciptomics and proteomics, including data-specific components in the model can be beneficial, as these data are different in, e.g., scale and measurement platform. For clinical and epidemiological studies, the multi-omics integration approaches, algorithmic and probabilistic two-way orthogonal partial least squares (O2PLS [5] resp. PO2PLS [7]), model the systematic difference between datasets. For small datasets, algorithmic approaches tend to overfit and a probabilistic approach is preferred [6]. However, these methods cannot directly be applied to data from cell lines, because they require that the datasets are measured on the same samples. To tackle these limitations, we proposed a new probabilistic approach, Probabilistic OPLS-DA (POPLS-DA), to deal with multiple omics data measured in cell lines by linking the datasets via genes instead of samples.

We present an integrated analysis of transcriptomics, proteomics, and FDA-approved drugs data using a novel probabilistic approach named POPLS-DA. We performed dose-response validation testing of all 53 compounds that were protective in the primary screening in at least one screening run. We applied POPLS-DA to integrate both transcriptomics and proteomics datasets and identified relevant genes and proteins that distinguished between the two experimental conditions (α-synuclein versus control protein overexpression). Based on these genes and proteins, a protein-protein interaction (PPI) network was constructed and the genes/proteins that are targeted by the drugs were identified. This approach highlighted genes in the network that are associated with α-synuclein overexpression and targeted by validated drugs, and therefore are potential targets for a novel therapy for synucleinopathies.

2. Results

2.1. Integration of transcriptome and proteome data

We applied our novel tool, POPLS-DA, to identify relevant genes/proteins by analyzing the pre-processed transcriptomics (x_t) and proteomics (x_p) datasets. Based on a scree plot of the eigenvalues of the row-wise concatenated x_t and x_p, we choose two joint and two specific components (Supplementary material). Based on a scree plot of the squared effect sizes per gene (given by the squared elements of $W\beta$), presented in Figure 1, 200 genes/proteins were retained for further analysis. A list of all genes/proteins and their weights are shown in the supplementary material.

Of the total variance in the transcriptomics and proteomics data, 21 and 18 percent, respectively, were explained by the integrative parts. Omics-specific variation accounted for around 27 percent of the total variation. The selected relevant genes/proteins appeared to be able to distinguish between the two experimental groups (Figure 1 and Supplementary material). We performed 400 permutations (see supplementary material) of the experimental group labels to test the null-hypothesis of no relation between the experimental groups and the genes/proteins. We found that, under the null hypothesis, the probability of achieving the same distinction (100% accuracy) as observed in the data was 0.0025 (one in 400 permutations).

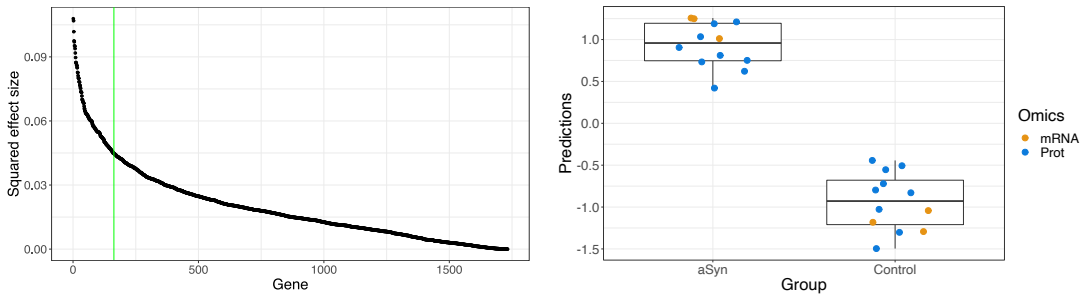

Figure 1: **Selection of relevant genes/proteins and their ability to discriminate.** The *left panel* shows the sorted squared effect size per gene given by the squared elements of $W\beta$. The 200 relevant genes/proteins correspond to the 'elbow', which is visually determined and approximately where the curve crosses the green vertical line. The *right panel* shows boxplots of score predictions based on the selected 200 relevant genes/proteins. The dots, representing individual samples, are added with a horizontal 'jitter' to reduce overlap. The transcriptomics and proteomics samples are colored orange and blue, respectively. Abbreviations: aSyn: α-synuclein, miRNA: transcriptome, Prot: proteome.

2.2. Validation of the protective efficacy of the FDA-approved drugs

Previously, we identified 53 FDA-approved drugs as being protective against αSyn-induced toxicity in a primary drug screening [14]. Now, using a dose-response analysis, 41 of these 53 compounds were validated as being protective. The maximum effective concentrations appeared to range from 2.5 μM to 80 μM. Figure 2 shows a schematic overview and results of the dose-response testing.

2.3. Omics and FDA data integration with String-DB

Figure 3 shows the protein-protein interaction network (αSyn-PPI) constructed using String-DB. The network had significantly more interactions than expected from an arbitrary subset of 200 genes. The network enrichment p-value was smaller than 10^{-15}. From the 41 drugs that were validated to reduce α-synuclein-induced toxicity [14], we found 27 targeting at least one gene in the αSyn-PPI network. The efficacy of the maximal effective dose of each of these compounds is shown in Table 1.

The *HMGCR*, *NQO1*, *HSPA5*, and *YBX1* genes were most targeted by the validated drugs, namely by 20, 18, 17, and 17 drugs, respectively.

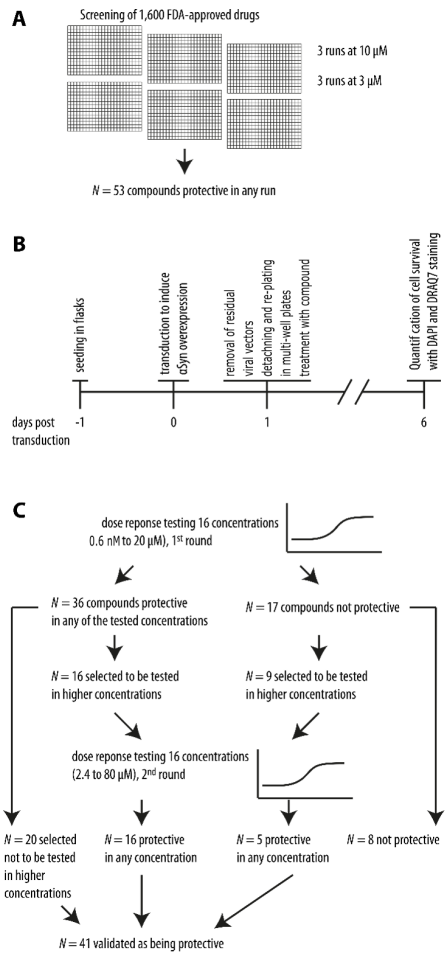

Figure 2: **Workflow of FDA-approved drug screening and dose-response validation.** 1,600 FDA-approved drugs were screened in two different concentrations (3 μM, 10 μM) in three runs each in an αSyn toxicity cell model. 53 compounds were identified as being protective in at least one run (A). Timeline of dose-response testing (B). In the first round, concentrations from 0.6 nM to 20 μM of each compound were investigated. 36 compounds could be confirmed as being protective. Of these, 20 already reached their maximal protective potential and were selected not to be tested in a higher concentration. The remaining 16 compounds were tested again in concentrations between 2.4 and 80 μM. All of these were confirmed as being protective in at least one concentration. Of the 17 compounds that were not protective in the first round dose-response testing, 8 were already toxic and considered not being protective. The remaining 9 were tested again in concentration from 2.4 to 80 μM. 5 of these compounds were protective in higher concentrations. In total, 41 could be confirmed as protective against αSyn induced toxicity (C).

Table 1 shows for each compound the number of direct neighbors it targets in the network. Dexibuprofen, a non-steroidal anti-inflammatory drug (NSAID), targets the highest number of direct neighbors (N=60), followed by Imatinib (N=47), which is acting on the tyrosine kinase enzyme and is involved in apoptosis. The most effective compounds were Telmisartan with an improvement of the cell survival of α-synuclein-

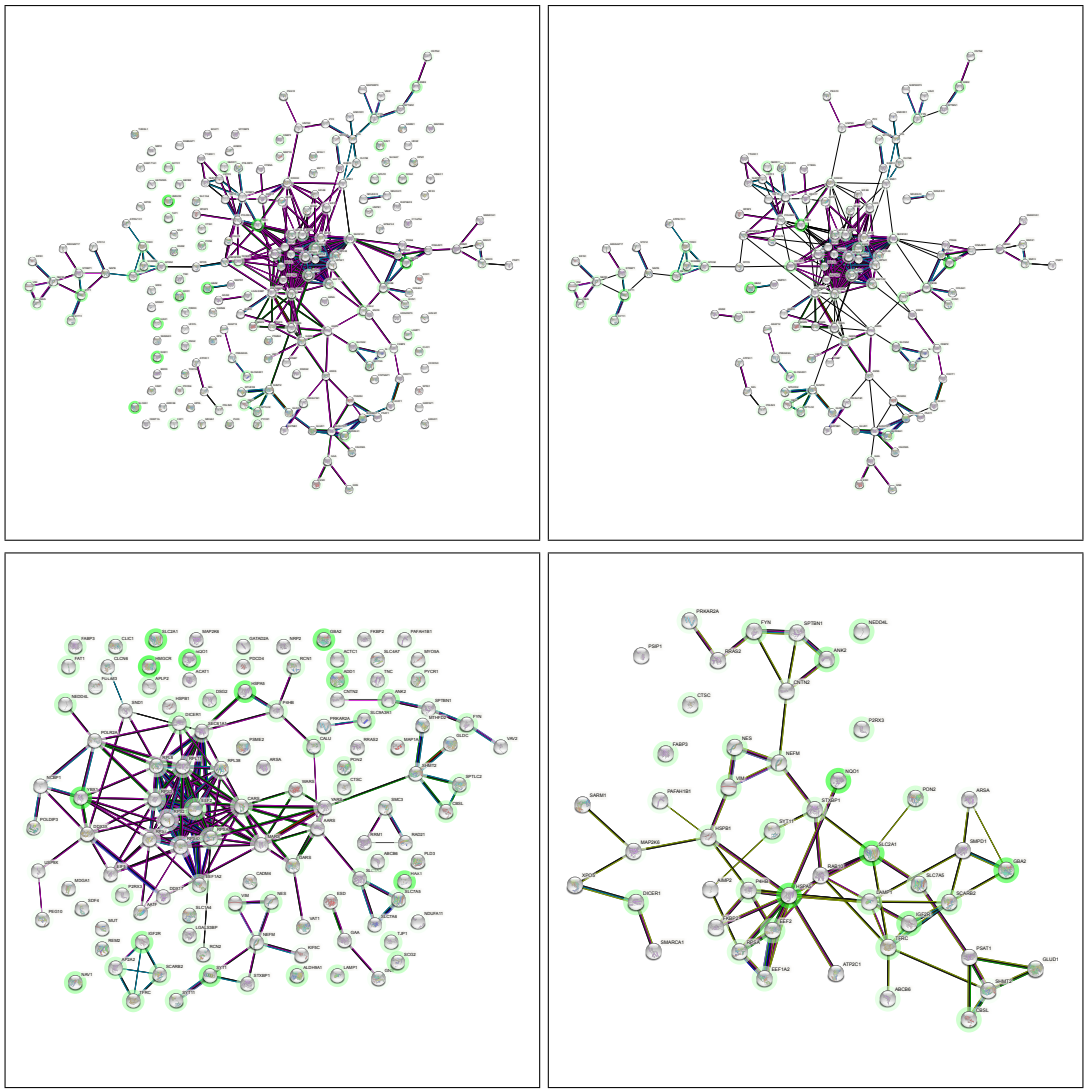

Figure 3: **String-DB and clustering analyses of the top 200 genes/proteins.** Figures are numbered from left to right, from top to bottom. In panel (a), a network of interactions between the top 200 genes/proteins (estimated with POPLS-DA) was constructed using String-DB. Each node is a gene, and a connection between genes indicates evidence for a biologically plausible link. Text-mining was excluded as an evidence source, and a medium confidence threshold was used. For genes that were (indirectly) targeted by a drug compound, a green 'halo' is drawn. The intensity of the green color is proportional to the number of drug compounds for which the gene was an (indirect) target. In panel (b), the interaction network from (a) was clustered using the MCL clustering algorithm from the String-DB website. The edges between the clusters are removed for visual aid. In panel (c), an interaction network is shown for a druggable subset of the top 200 genes/proteins, consisting of 116 genes that were (indirectly) targeted by an FDA-approved drug compound. In panel (d), an interaction network of top genes in the "Parkinson's disease" DisGeNet term was constructed using String-DB. Text mining as evidence was included here.

overexpressing cells by 33.16% at a concentration of 5 μM, followed by risperidone which improved cell survival by 28.58% at 10 μM.

Drug name	**Direct neighbors**	**Most effective concentration**	**Cell survival**	**P-value**
Dexibuprofen	60	80 μM	118.29%	<0.0001
Imatinib	47	20 μM	126.04%	<0.0001
Amiodarone	44	80 mM	108.05%	0.002
Ripseridone	41	10 μM	128.58%	<0.0001
Bepridil	36	80 μM	115.89%	<0.0001
Astemizole	26	1.25 μM	106.49%	0.02
Telmisartan	24	5 μM	133.16%	<0.0001
Amlodipine	21	10 μM	116.42%	<0.0001
Benazepril	18	80 μM	117.96%	<0.0001
Repserpine	18	80 μM	107.64%	0.01
Dipyridamole	16	40 μM	118.09%	<0.0001
Trifluopromazine	16	40 μM	114.30%	<0.0001
Nefazodone	15	40 μM	115.08%	<0.0001
Trimipramine	15	20 μM	125.54%	<0.0001
Flunarizine	14	20 μM	104.38%	0.03
Quinacrine	13	5 μM	119.07%	0.008
Pentoxyverine	10	1.25 μM	109.69%	0.002
Dicyclomine	8	20 μM	120.15%	<0.0001
Doxazosin	8	40 μM	116.12%	<0.0001
Ajmaline	7	20 μM	108.94%	<0.0001
Ifenprodil	7	2.5 μM	116.74%	0.0003
Lomerizine	7	80 μM	110.35%	<0.0001
Pentamidine	6	2.5 μM	119.86%	0.003
Guanfacine	5	5 μM	117.01%	<0.0001
Dyclonine	4	5 μM	125.69%	<0.0001
Tropisetron	3	20 μM	131.63%	<0.0001
Clemastine	1	10 μM	117.52%	0.0004

Table 1: **FDA-approved drugs protective against α-synuclein-induced toxicity with direct neighbors of the top 200 genes.** For each drug, the number of direct interactors is shown. The most effective concentration is the concentration at which each compound had the strongest protective effect. Cell survival is expressed as percentage compared to αSyn-overexpressing cells that were treated with DMSO as control at the maximum effective concentration. P-Values of testing the dose-reponse were calculated with an ANOVA with Dunnett's post-hoc test for multiple comparisons.

The network was split into sub-clusters using the MCL clustering algorithm on the αSyn-PPI network, see Figure 3. Several clusters consist of genes that are involved in neurogenerative disorders. The largest cluster is a highly connected hub of 22 genes, primarily ribosomal. Ribosomal proteins have been reported to mediate neurodegenera-

tion in PD [27]. Another cluster of five genes includes the top two genes, *MTHFD2* and *CBS*. Overexpression of the *CBS* (cystathionine-beta-synthase) gene was protective in a PD toxin model in rats [44]. The network also contains clusters with genes involved in neurological development, e.g. *TFRC* (transferrin receptor), and neuronal adhesion and growth, e.g. *TNC* (tenascin C). Further, four genes encoding intermediate filaments, such as *NES* (nestin), are clustered together. They are primarily responsible for cell adhesion [30]. Furthermore, a cluster of three genes from the solute carrier family was found (*SLC3A2*, *SLC7A5*, and *SLC7A6*). These genes may play a role in the transport across the blood-brain barrier.

2.4. Enrichment analyses with GO and DisGeNet

In Table 2, the most significant pathways enriched in the top 200 genes are shown. The GO terms that contained "extracellular" and "cytoplasm" were the most significant. Genes with these annotations are involved in vesicular transport and signaling [19]. Nine of the ten terms were from the "cellular compound" GO category.

Term	**Ontology**	**Genes in set**	**FDR**
cytoplasm	CC	178	1.28e-19
extracellular exosome	CC	71	4.67e-17
extracellular vesicle	CC	71	5.51e-16
extracellular organelle	CC	71	5.51e-16
intracellular anatomical structure	CC	190	7.62e-12
cell junction	CC	61	7.62e-12
vesicle	CC	88	2.30e-11
extracellular space	CC	81	7.09e-11
protein binding	CC	179	7.65e-11
cytosol	MF	100	1.48e-10

Table 2: **Gene ontology enrichment analysis of the top 200 genes.** Out of 1732 genes/proteins, the 200 genes/proteins with the highest weight (estimated using POPLS-DA) were used to perform GO enrichment analysis. All three GO ontologies were considered: cellular compound (CC), biological process (BP), and molecular function (MF). For each term, a p-value was computed using Fisher's exact test with the whole genome as background and corrected for multiple testing. The ten most significant terms are shown, ranked by their GO enrichment false discovery rate (FDR) adjusted p-value. The column "Genes in set" shows how many of the top 200 genes/proteins were annotated with the respective term.

Table 3 shows the five diseases for which the most significant enrichment was found in the top 200 genes. The second term was "Parkinson's disease" and contained 48 genes. Figure 3 shows the PPI network for genes/proteins in this disease term. Note that the *HSPA5* gene, which was one of the most targeted by the FDA drug compounds, is within this network.

Finally, enrichment analyses were performed for the subset of 116 genes/proteins that were direct neighbors of genes targeted by at least one of the drugs. Figure 3 shows the protein-protein interaction network for this subset. The corresponding network

Term	Genes in set	FDR
Neoplasms	142	2.84e-07
Parkinson's Disease	48	1.94e-06
Carcinogenesis	100	3.36e-06
Malignant neoplasm of prostate	75	2.54e-05
Neoplasm Metastasis	98	2.89e-05

Table 3: **Top five disease enrichment clusters based on DisGeNet.** The top 200 relevant genes/proteins were analyzed using DisGeNet, a disease-gene database. The BEFREE database was used. For each term, a p-value was calculated using a Fisher exact test and FDR adjusted. The terms are ranked by this adjusted p-value.

enrichment p-value was smaller than 10^{-10}. Also, the most significant GO terms agreed with the GO analysis of the total set of 200 genes/proteins (Supplementary material).

3. Discussion

In this study, we integrated data from LUHMES cells overexpressing α-synuclein and those overexpressing GFP as a control protein. Transcriptomics and proteomics data were mapped to common gene identifiers, and jointly analyzed with POPLS-DA. We found 200 relevant genes/proteins based on an elbow plot of the squared effect size of each gene (Figure 1). Using a dose-response approach, 41 FDA-approved drugs were validated to be protective against α-synuclein-induced toxicity. A protein-protein interaction network was constructed based on the 200 relevant genes/proteins. The data from the FDA-approved drug validation were integrated by identifying direct neighbors of their drug targets in the network. These genes/proteins appeared to perfectly discriminate between the two experimental groups. Interactome analyses revealed several genes and pathways linked to PD, the most common synucleinopathy.

The drugs that targeted the highest number of direct neighbors were Dexibuprofen and Imatinib. Both drugs were previously found to lower the risk of PD in epidemiological studies [33, 20]. Interestingly, independent of the screening, we previously investigated the neuroprotective potential of imatinib as a macroautophagy-stimulating drug. Despite the protective effect, imatinib in our model led to a reduction of the expression of the dopaminergic markers tyrosine hydroxylase (TH) and dopamine transport (DAT) [13]. Therefore, further investigation would be necessary before using this drug in PD patients. However, other drugs we identified could be readily investigated in PD. Telmisartan, an AT1-blocker, which is used for arterial hypertension, was the most effective drug with an improvement of cell survival >30% at a comparably low concentration of 5μM. Since there are no major restrictions in PD other than a blood pressure lowering effect, telmisartan could be a promising candidate for a neuroprotective therapy. Similarly, also benazepril, an ACE-inhibitor and also blood-pressure lowering drug could be investigated with similar caution. Interestingly, AT1-blockers and ACE inhibitors have been shown to reduce the risk of falls in PD, and this association was not related to blood pressure [21]. Furthermore, the use of AT1-blockers and ACE inhibitors as

potential neuroprotective drugs was previously discussed, because these drugs were protective in PD animal models, most likely due to an anti-oxidative effect [29]. In line with that, our data further emphasize that it would be promising to further investigate AT1-blockers and ACE-inhibitors in PD.

Other drugs, including risperidone and trifluopromazine, are antipsychotic drugs with an anti-dopaminergic effect and are therefore contraindicated in PD and can even lead to drug-induced parkinsonism [23]. Also, Flunarizine, a drug used in migraine prophylaxis can lead to drug-induced parkinsonism [41, 22] and would therefore not be the first choice in PD.

The genes that were targeted by the majority of the drugs were *HMGCR*, *NQO1*, *HSPA5*, and *YBX1*. Note that *HMGCR* (3-Hydroxy-3-Methylglutaryl-CoA reductase) is the main target of statins. *NQO1* (NAD(P)H quinone dehydrogenase 1) has multiple functions and has been linked to Alzheimer's disease [37]. The *HSPA5* gene (heat shock protein family A (Hsp70) member 5) encodes a heat shock protein of the HSP70 family. It is found to be involved in maintaining the correct folding behavior of proteins [25]. HSPA5, also known as glucose-regulated protein 78 (GRP78), is located on the membrane of the endoplasmic reticulum and is essential for the unfolded protein response [17]. Furthermore, it was previously shown that overexpression of this protein reduced α-synuclein toxicity to dopaminergic cells in a rat PD model [10]. Interestingly, a previous study showed that GRP78 levels were reduced in the temporal cortex and cingulate gyrus of PD patients compared to healthy controls but markedly increased in PD patients with dementia and patients with dementia with Lewy bodies [1]. In line with that, our data that show an interaction of drugs that protect against α-Syn with this protein support that GRP78 plays an important role in α-Syn toxicity and deliver further evidence that GRP78 could be an promising target for neuroprotective therapies of PD. The *YBX1* gene (Y-box binding protein 1) is a cold shock domain protein, predominantly localized to neurons and essential for brain development [3]. Furthermore, several clusters were observed in the network involving ribosomal proteins and intermediate filaments. Enrichment analyses identified several terms involving "extracellular" and "vesicles".

Using DisGeNet, the second most significantly enriched disease term was "Parkinson's disease", with 48 genes from our list of relevant genes/proteins. The other disease terms involved neoplasms and carcinogenesis. Note that previous studies suggested shared molecular mechanisms between PD and cancer [40, 4, 9].

In a previous study on these data [15], the authors identified significant deregulation of synaptic proteins, in particular, *STXBP1*, *STX1B*, *SYT1*, and *FYN*. In our POPLS-DA integration, we found three of these genes: *STXBP1*, *SYT1*, and *FYN*, ranked 16, 18, and 72, respectively. We identified two additional synaptic genes in the top 200, namely *STX12*, encoding syntaxin-12, as well as *SYT11*, encoding synaptotagmin-11. Single nucleotide polymorphisms in *SYT11* have been linked to PD in a previously performed GWAS [18]. Furthermore, a previous study showed that synaptotagmin-11 plays a role in Parkin-related parkinsonism. Additionally, it was previously reported that lysosome related proteins, *SCARB2*, *CTSB*, *SMPD1*, were differentially regulated upon

overexpression of α-synuclein in LUHMES cells [15]. From our top genes/proteins, 22 were annotated with the lysosome GO term. Among these 22 genes are the previously found *SCARB2* at position 7 in our top genes/proteins, as well as *SLC7A5* and *SLC3A2* at position 5 and 6, respectively. The authors pointed to lysosome membrane protein 2 expression (encoded by *SCARB2*) as a therapeutic target to increase lysosomal glucocerebrosidase activity to promote α-synuclein clearance. Note that one of our clusters in Figure 3 consists of genes encoding lysosome related proteins (i.e., *SCARB2*, *AP2A2*, *ATP6V1C1*, *IGF2R*, and *TFRC*) that are also direct neighbors of the FDA-approved drugs.

The data used in this study were acquired from a human (LUHMES) cell line, where each dataset was measured in a different sample. Most data integration methods require multiple datasets measured in the same samples to calculate correlations across omics datasets from which joint components can be derived. These methods are sometimes referred to as horizontal integration approaches, where the data are laid out next to each other with the samples across the rows. In the present paper, we proposed a framework for vertical integration, where different groups of samples are laid down below each other with the same (type of) variables across the columns. A drawback of our approach is that only genes/proteins with observations in both datasets can be analyzed.

We utilized four bioinformatics databases in our analysis: String-DB, GeneBank, GO, and DisGeNet. Several databases are available with information on the interaction between genes and other molecular variables (interactomes) [34]. We chose String-DB since it was found to perform well with regard to disease prediction [28]. However, they also showed that other databases might comprise complementary information. Currently, to the best of our knowledge, no method can utilize all information across different interactomes. Information bias is another unsolved problem when using these databases: less studied diseases have less coverage in these databases [12]. Since synucleinopathies are less common and less studied (compared to, e.g., neoplasms), the bioinformatics interpretation of our results are likely to be biased towards well-studied disease. Indeed, in DisGeNet, PD has only 2078 annotated genes, while neoplasms have more than 10000 genes. This could explain that our enrichment analysis based on DisGeNet (Table 3) showed many genes annotated with "neoplasms".

The POPLS-DA method can integrate more than two datasets to distinguish between experimental conditions in cell lines if the variables in these datasets can be mapped to the same unit, i.e. genes. For example, SNPs from a GWAS study can be integrated by mapping them to genes. Here the distance of a SNP to a gene or its transcription factor binding site [16] can be used. For methylation, CpG sites are commonly mapped to genes based on their distance to a transcription start site [2]. Since, typically, multiple SNPs and CpG sites are mapped to the same gene, an aggregation step is needed to obtain one measurement per gene. For example, the first principal component of the SNPs or CpG sites corresponding to a gene can be taken as the final aggregated measurement.

The data from the FDA-approved drugs were integrated and projected onto the interaction network of the 200 relevant genes obtained from the transcriptomics and proteomics data analyses. An alternative approach is to simultaneously analyze the

transcriptomics and proteomics data together with data on the validated compounds, i.e. drug-augmented data integration. It has been shown that including such information via penalties into data integration analysis yields robust results [11]. Such a joint model can pinpoint genes that can discriminate the two groups while targeted by the compounds, improving the clinical relevance of the list of top genes.

Conclusion. We validated 41 FDA-approved drugs to be protective against α-synuclein-induced toxicity in dose-response analyses. We applied a novel data integration approach to combine these drug data with experimental multi-omics data. We identified a set of 200 relevant genes/proteins that discriminated between samples overexpressing α-synuclein and controls, as well their drug targets. Some of the drugs (e.g. an ACE-inhibitor and an AT1-blocker) could be readily investigated in PD patients. These findings can potentially be used to develop therapeutic targets for Parkinson's disease, multiple system atrophy, and other synucleinopathies.

4. Materials and Methods

4.1. Processing samples and measuring data

The screening of 1600 FDA-approved drugs and the acquisition of proteome and transcriptome data are described in detail elsewhere [14, 15] and summarized here in brief.

FDA-approved drug compound validation. For the dose-response testing, LUHMES cells were plated in flasks coated with poly-L-lysine (0.1 mg/ml) and fibronectin (5 μg/ml) in growth medium (DMEM/F12 with 1% N2 supplement and 0.04 μl/ml basic fibroblast growth factor). After 24h, the medium was changed to differentiation medium (DMEM/F12 with 1% N2-supplement, 1 μg/ml tetracycline, 0.49 μg/ml N6,2'-O-dibutyryladenosine 3',5'-cyclic monophosphate, 2 ng/ml glial derived neurotrophic factor). After another 24 hours, the cells were transduced with adenoviral vectors leading to overexpression of human wild-type α-synuclein. On the day after transduction, viral vectors were washed away, and the cells were replated on 384-well multi-well plates and treated with the different drugs solved in dimethyl sulfoxide. For validation, compounds were tested in concentrations between 0.6 nM to 20 μM. Compounds that were not protective or did not reach a protective plateau were further tested in higher concentrations up to 80 μM. The dose-response data were analyzed using an ANOVA with Dunnett's post hoc test to correct for multiple comparison with GraphPad Prism version 9.4 for Windows 64-bit (GraphPad Software, San Diego, CA, USA).

Transcriptome and proteome data acquisition and processing. For the transcriptome and proteome data, cells were plated directly in differentiation medium and transduced with adenoviral vectors leading to overexpression of either human wild-type α-synuclein or GFP, respectively. After 24 hours, the remaining viral vectors were removed. Samples were collected four days after transduction. Transcriptome data were measured with Illumina HumanHT-12_V3 bead chips (Illumina, San Diego, CA, USA). Proteome data

were measured by liquid chromatography-mass spectrometry (LC-MS/MS). The transcription probes were filtered based on the Illumina detection p-value, where probes were removed when all samples had a p-value above 0.05. A variance filter was applied where probes were removed if the inter-quantile range (IQR) of a probe did not exceed the 0.5-th quantile of all IQRs. Probes that could not be mapped to an Entrez identifier were removed.

After pre-processing and filtering, the transcriptomics and proteomics datasets contained 15660 transcription probes and 2577 proteins that could be mapped to Entrez identifiers. 2292 of these identifiers were present in both datasets. We excluded the *SNCA* measurements from the data analysis.

For the transcriptomics data, we analyzed three samples from LUHMES cells overexpressing α-synuclein, and three samples from LUHMES cells expressing GFP as a control protein. For proteomics, the numbers were equal to nine. The correlation of the transcripts with the α-synuclein overexpressing cells versus the control cells was taken as a reference; if this correlation was negative, the corresponding protein feature was multiplied by minus one. The measurements for both datasets were scaled to have zero mean and unit variance.

4.2. Probabilistic data integration with POPLS-DA

For the joint analysis of the transcriptomics and proteomics datasets, a novel statistical data integration method, probabilistic orthogonal partial least squares discriminant analysis (POPLS-DA) was applied. Details can be found in the supplementary materials.

Briefly, let x_t and x_p be random vectors representing the 2292 transcriptomics and proteomics data with the same Entrez identifier, respectively. Similarly, let y_t (for transcriptomics) and y_p (for proteomics) be univariate random variables encoding the two experimental conditions (one for cells overexpressing α-synuclein, and zero for the controls). In the POPLS-DA model, the group status and omics data are linked via latent variables u_t and u_p of size r, with r much smaller than the data dimensions. These latent variables represent joint parts of the two omics datasets that are correlated with the group status. To model omics variables that contribute to the variation in the omics dataset but do not play a role in distinguishing between the two experimental groups, specific components v_t and v_p were added to the model. The numbers of components here are denoted by r_t and r_p, respectively. The number of joint and specific components, r, r_t, and r_p have to be specified and can be based on scree plots [26] of the eigenvalues of the stacked datasets. Residual variation was modeled by including noise terms e_t and e_p for the transcriptomics and proteomics datasets, and ϵ_t and ϵ_p for the random variable representing the groups. All random variables are assumed to be normally distributed with a mean of zero. The mathematical model for POPLS-DA can be written as

$$\begin{aligned} x_t &= u_t W^{\mathrm{T}} + v_t P_t^{\mathrm{T}} + e_t, & x_p &= u_p W^{\mathrm{T}} + v_p P_p^{\mathrm{T}} + e_p, \\ y_t &= u_t \beta + \epsilon_t, & y_p &= u_p \beta + \epsilon_p. \end{aligned} \tag{1}$$

The matrix W contains the joint loadings, and the matrices P_t and P_p contain the specific loadings for each gene in each component. The vector β contains the regression

coefficients of y_t on the joint components u_t. Note that these regression coefficients are the same for y_p on the joint components u_p. The vector $W\beta$ represents effect sizes for how well each gene, represented by its transcript and protein, discriminates between the two experimental conditions.

The parameters of POPLS-DA were estimated with maximum likelihood. See the supplementary materials for more details. To select a set of relevant genes/proteins, the obtained effect sizes $W\beta$ were ranked and plotted. The elbow criterion [26] was used to determine the threshold between relevant and irrelevant genes/proteins.

Our R implementation of the POPLS-DA method is freely available online under repository: github.com/selbouhaddani/POPLSDA.

4.3. FDA-approved drugs analysis

The relevant genes/proteins identified by POPLS-DA were further analyzed using databases on protein-protein interactions (PPI) and drug targets, as well as gene ontology and gene-disease association databases.

The PPI network of the relevant genes/proteins (in short, αSyn-PPI) was estimated using the String-DB [39] website and the `STRINGdb` package in R [32]. In the String-DB analysis, the text mining evidence was excluded; the network was based only on curated, computational, and experimental data sources. A medium evidence threshold (0.4) was used to define an interaction (edge in the network).

To integrate the FDA-approved drug screening data with the obtained αSyn-PPI network, we first identified drugs that significantly reduced the toxicity of α-synuclein aggregation. We used DrugBank [43] to retrieve the list of genes targeted by these drugs. Using the `string_db$get_neighbors` function, the direct neighbors of these targeted genes in the String-DB database were added to the gene lists for each compound. Then, the number of compounds targeting each gene in the α-Syn-PPI network was determined. Nodes (genes and proteins) in the network were colored according to this number using the "payload" function on the String-DB website.

To identify sub-clusters of genes/proteins in the αSyn-PPI network, the Markov cluster algorithm (MCL) [8] on the String-DB website was used. The inflation parameter was set to two.

Finally, gene ontology (GO) pathway and disease enrichment analyses were performed. For GO, the `goana` function from the `limma` package [36] in R was used, with the whole human genome as background. For disease enrichment analysis, DisGeNet [31], a disease-gene association database implemented in the `disgenet2r` R package, was used. Here, the "BEFREE" text mining database was used for the disease-gene enrichment score as it contains more annotations regarding MSA and PD. The integration methodology is presented as a workflow in Figure 4.

Acknowledgements

All authors were funded by ERA-Net E-Rare JTC 2018 (MSA-omics) [40-44000-98-2006/ 90030376507]. JD was supported by European Union's Horizon 2020 (IMforFUTURE) [721815] and the EU funded Cost action DYNALIFE [CA21169]. GH

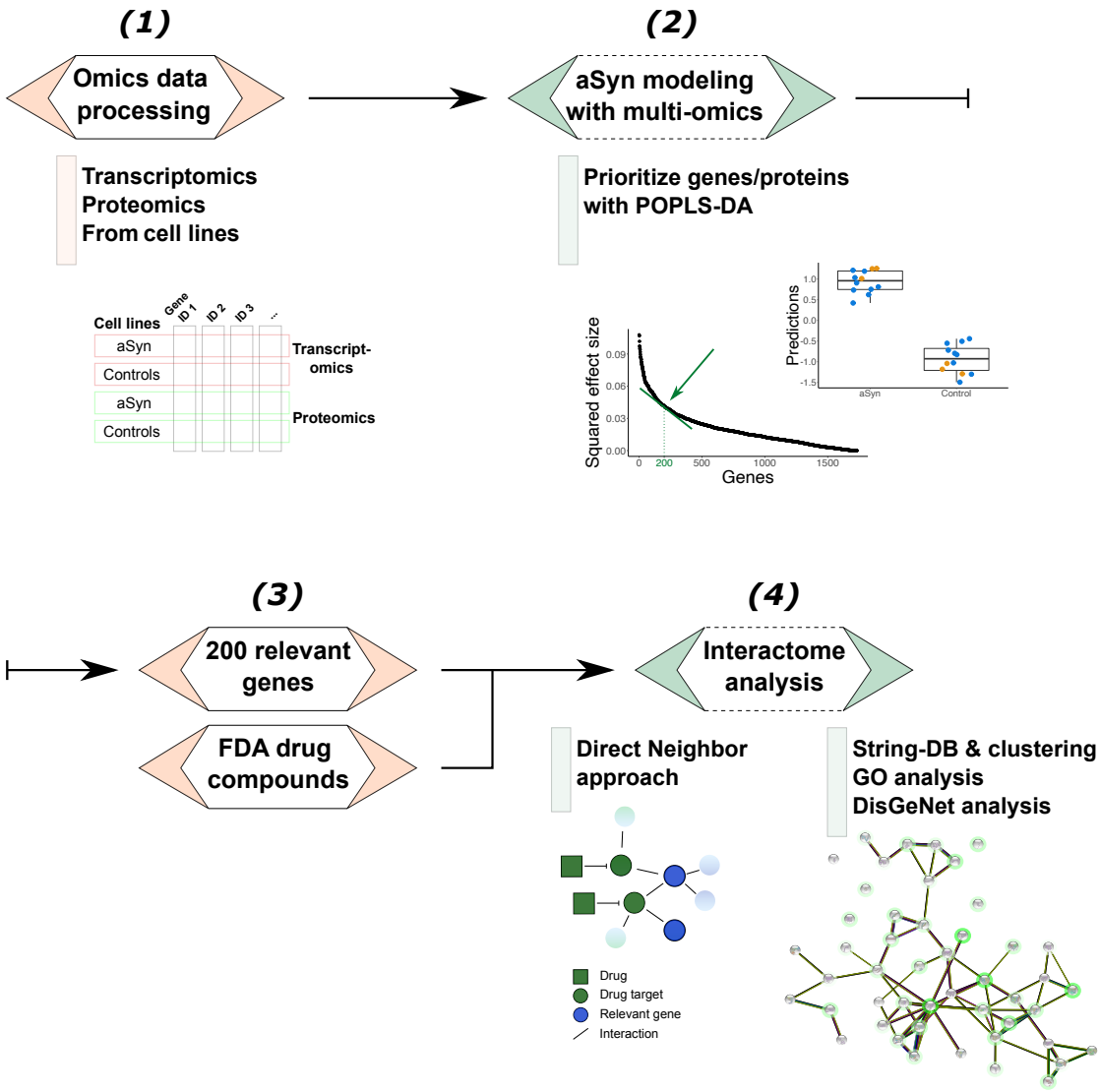

Figure 4: **Workflow of the multi-omics integration approach.** (1) To link the two datasets, the transcripts and proteins were mapped to their Entrez identifier, referred to as genes/proteins. (2) To identify the relevant genes/proteins that discriminate between the two experimental groups (α-synuclein overexpression vs control), POPLS-DA was applied. (3) The prioritized genes/proteins were integrated with the FDA-approved drug screening data using a direct neighbor approach. As an illustration, in the figure under "Direct Neighbor approach" are two drug target genes (green circles), which are neighbors of genes identified in (2) (blue circles). (4) Bioinformatics analyses: interactome, GO and DisGeNet enrichment analyses analyses. Protein-protein interaction networks were built using String-DB where each node is a gene. The genes that were direct neighbors of a drug target were highlighted with a green halo, where its color intensity is proportional to the number of direct neighbors. Abbreviations: aSyn: α-synuclein, GO: gene ontology , FDA: Food and Drug Administration.

was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the framework of the Munich Clus-

ter for Systems Neurology (EXC 2145 SyNergy) [390857198] and within the Hannover Cluster RESIST (EXC 2155) [39087428], DFG [HO2402/18-1] MSAomics, the Bavarian Ministry for Education, Culture, Science and Art (ForIPS) [8810001412], Niedersächsisches Ministerium für Wissenschaft und Kunst [ZN3440.TP]: REBIRTH – Forschungszentrum für translationale regenerative Medizin, VolkswagenStiftung (Niedersächsisches Vorab), Petermax-Müller Foundation (Etiology and Therapy of Synucleinopathies and Tauopathies). MH and GH received funding from ParkinsonFonds Deutschland (Hypothesis-free compound screening in a new human neuronal model of wild type alpha-synuclein-induced cell death).

Conflict of interest

The authors report no conflicts of interest.

References

[1] Jean-Ha Baek, Dejan Mamula, Beata Tingstam, Marcela Pereira, Yachao He, and Per Svenningsson. GRP78 Level Is Altered in the Brain, but Not in Plasma or Cerebrospinal Fluid in Parkinson's Disease Patients. *Front. Neurosci.*, 13:697, 2019.

[2] Marina Bibikova, Bret Barnes, Chan Tsan, Vincent Ho, Brandy Klotzle, Jennie M. Le, David Delano, Lu Zhang, Gary P. Schroth, Kevin L. Gunderson, Jian-Bing Bing Fan, and Richard Shen. High density DNA methylation array with single CpG site resolution. *Genomics*, 98(4):288–295, oct 2011.

[3] N. V. Bobkova, D. N. Lyabin, N. I. Medvinskaya, A. N. Samokhin, P. V. Nekrasov, I. V. Nesterova, I. Y. Aleksandrova, O. G. Tatarnikova, A. G. Bobylev, I. M. Vikhlyantsev, M. S. Kukharsky, A. A. Ustyugov, D. N. Polyakov, I. A. Eliseeva, D. A. Kretov, S. G. Guryanov, and L. P. Ovchinnikov. The Y-Box Binding Protein 1 Suppresses Alzheimer's Disease Progression in Two Animal Models. *PLoS One*, 10(9):e0138867, sep 2015.

[4] Maria Ejma, Natalia Madetko, Anna Brzecka, Konstanty Guranski, Piotr Alster, Marta Misiuk-Hojło, Siva G. Somasundaram, Cecil E. Kirkland, and Gjumrakch Aliev. The Links between Parkinson's Disease and Cancer. *Biomedicines*, 8(10):416, oct 2020.

[5] Said el Bouhaddani, Jeanine Houwing-Duistermaat, Perttu Salo, Markus Perola, Geurt Jongbloed, and Hae-Won Uh. Evaluation of O2PLS in Omics data integration. *BMC Bioinformatics*, 17(S2):S11, dec 2016.

[6] Said el Bouhaddani, Hae-Won Uh, Geurt Jongbloed, Caroline Hayward, Lucija Klarić, Szymon M. Kiełbasa, and Jeanine Houwing-Duistermaat. Integrating omics datasets with the OmicsPLS package. *BMC Bioinformatics*, 19(1):371, dec 2018.

[7] Said el Bouhaddani, Hae-Won Uh, Geurt Jongbloed, and Jeanine Houwing-Duistermaat. Statistical integration of heterogeneous omics data: Probabilistic two-way partial least squares (PO2PLS). *J. R. Stat. Soc. Ser. C (Applied Stat.*, aug 2022.

[8] A. J. Enright, S. Van Dongen, and C. A. Ouzounis. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, 30(7):1575–84, apr 2002.

[9] Panagiota S. Filippou and Tiago F. Outeiro. Cancer and Parkinson's Disease: Common Targets, Emerging Hopes. *Mov. Disord.*, 36(2):340–346, feb 2021.

[10] Marina S Gorbatyuk, Arseniy Shabashvili, Weijun Chen, Craig Meyers, Layla F Sullivan, Max Salganik, Jonathan H Lin, Alfred S Lewin, Nicholas Muzyczka, and Oleg S Gorbatyuk. Glucose regulated protein 78 diminishes α-synuclein neurotoxicity in a rat model of Parkinson disease. *Mol. Ther.*, 20(7):1327–37, jul 2012.

[11] Zhujie Gu, Said el Bouhaddani, Jiayi Pei, Jeanine Houwing-Duistermaat, and Hae-Won Uh. Statistical integration of two omics datasets using GO2PLS. *BMC Bioinformatics*, 22(1):131, dec 2021.

[12] Winston A. Haynes, Aurelie Tomczak, and Purvesh Khatri. Gene annotation bias impedes biomedical research. *Sci. Rep.*, 8(1):1362, dec 2018.

[13] Matthias Höllerhage, Natascha Fussi, Thomas W Rösler, Wolfgang Wurst, Christian Behrends, and Günter U Höglinger. Multiple molecular pathways stimulating macroautophagy protect from alpha-synuclein-induced toxicity in human neurons. *Neuropharmacology*, 149:13–26, may 2019.

[14] Matthias Höllerhage, Claudia Moebius, Johannes Melms, Wei-Hua Chiu, Joachim N Goebel, Tasnim Chakroun, Thomas Koeglsperger, Wolfgang H Oertel, Thomas W Rösler, Marc Bickle, and Günter U Höglinger. Protective efficacy of phosphodiesterase-1 inhibition against alpha-synuclein toxicity revealed by compound screening in LUHMES cells. *Sci. Rep.*, 7(1):11469, dec 2017.

[15] Matthias Höllerhage, Markus Stepath, Michael Kohl, Kathy Pfeiffer, Oscar Wing ho Chua, Linghan Duan, Franziska Hopfner, Martin Eisenacher, Katrin Marcus, and Günter U. Höglinger. Transcriptome and Proteome Analysis in LUHMES Cells Overexpressing Alpha-Synuclein. *Front. Neurol.*, 13(April):1–19, apr 2022.

[16] Xinli Hu, Hyun Kim, Eli Stahl, Robert Plenge, Mark Daly, and Soumya Raychaudhuri. Integrating Autoimmune Risk Loci with Gene-Expression Data Identifies Specific Pathogenic Immune Cell Subsets. *Am. J. Hum. Genet.*, 89(4):496–506, oct 2011.

[17] Ibrahim M Ibrahim, Doaa H Abdelmalek, and Abdo A Elfiky. GRP78: A cell's response to stress. *Life Sci.*, 226:156–163, jun 2019.

[18] International Parkinson Disease Genomics Consortium, Michael A Nalls, Vincent Plagnol, Dena G Hernandez, Manu Sharma, Una-Marie Sheerin, Mohamad Saad, J Simón-Sánchez, Claudia Schulte, Suzanne Lesage, Sigurlaug Sveinbjörnsdóttir, Kári Stefánsson, Maria Martinez, John Hardy, Peter Heutink, Alexis Brice, Thomas Gasser, Andrew B Singleton, and Nicholas W Wood. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet (London, England)*, 377(9766):641–9, feb 2011.

[19] Chen-Yuan Kao and Eleftherios T. Papoutsakis. Extracellular vesicles: exosomes, microparticles, their parts, and their targets to enable their biomanufacturing and clinical applications. *Curr. Opin. Biotechnol.*, 60:89–98, dec 2019.

[20] Manish Kumar, Ritu Kulshrestha, Nirmal Singh, and Amteshwar Singh Jaggi. Expanding spectrum of anticancer drug, imatinib, in the disorders affecting brain and spinal cord. *Pharmacol. Res.*, 143:86–96, may 2019.

[21] Alice Laudisio, Maria Rita Lo Monaco, Maria Caterina Silveri, Anna Rita Bentivoglio, Davide L Vetrano, Maria Stella Pisciotta, Vincenzo Brandi, Roberto Bernabei, and Giuseppe Zuccalà. Use of ACE-inhibitors and falls in patients with Parkinson's disease. *Gait Posture*, 54:39–44, may 2017.

[22] Wei Lin, Cheng-Li Lin, Chung Y Hsu, and Cheng-Yu Wei. Flunarizine Induced Parkinsonism in Migraine Group: A Nationwide Population-Based Study. *Front. Pharmacol.*, 10:1495, 2019.

[23] José Luis López-Sendón, María Angeles Mena, and Justo García de Yébenes. Drug-induced parkinsonism in the elderly: incidence, management and prevention. *Drugs Aging*, 29(2):105–18, feb 2012.

[24] Julie Lotharius. Progressive Degeneration of Human Mesencephalic Neuron-Derived Cells Triggered by Dopamine-Dependent Oxidative Stress Is Dependent on the Mixed-Lineage Kinase Pathway. *J. Neurosci.*, 25(27):6329–6342, jul 2005.

[25] Miles S. Lyon and Carol Milligan. Extracellular heat shock proteins in neurodegenerative diseases: New perspectives. *Neurosci. Lett.*, 711:134462, oct 2019.

[26] Kantilal Varichand Mardia, John T Kent, and John M Bibby. *Multivariate analysis*. Academic Press, 1979.

[27] Ian Martin, Jungwoo Wren Kim, Byoung Dae Lee, Ho Chul Kang, Jin-Chong Xu, Hao Jia, Jeannette Stankowski, Min-Sik Kim, Jun Zhong, Manoj Kumar, Shaida A. Andrabi, Yulan Xiong, Dennis W. Dickson, Zbigniew K. Wszolek, Akhilesh Pandey, Ted M. Dawson, and Valina L. Dawson. Ribosomal Protein s15 Phosphorylation Mediates LRRK2 Neurodegeneration in Parkinson's Disease. *Cell*, 157(2):472–485, apr 2014.

[28] Ettore Mosca, Matteo Bersanelli, Tommaso Matteuzzi, Noemi Di Nanni, Gastone Castellani, Luciano Milanesi, and Daniel Remondini. Characterization and comparison of gene-centered human interactomes. *Brief. Bioinform.*, 00(October 2020):1–16, 2021.

[29] Santiago Perez-Lloret, Matilde Otero-Losada, Jorge E Toblli, and Francisco Capani. Renin-angiotensin system as a potential target for new therapeutic approaches in Parkinson's disease. *Expert Opin. Investig. Drugs*, 26(10):1163–1173, oct 2017.

[30] Rodolphe Perrot and Joel Eyer. *Intermediate Filaments in Neurodegenerative Diseases*. InTech, may 2013.

[31] Janet Piñero, Juan Manuel Ramírez-Anguita, Josep Saüch-Pitarch, Francesco Ronzano, Emilio Centeno, Ferran Sanz, and Laura I. Furlong. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.*, 48(D1):D845–D855, jan 2020.

[32] R Core Team. R: A Language and Environment for Statistical Computing, 2014.

[33] Karen Rees, Rebecca Stowe, Smitaa Patel, Natalie Ives, Kieran Breen, Carl E Clarke, and Yoav Ben-Shlomo. Non-steroidal anti-inflammatory drugs as disease-modifying agents for Parkinson's disease: evidence from observational studies. *Cochrane Database Syst. Rev.*, (11), nov 2011.

[34] Daniel J Rigden and Xosé M. Fernández. The 2022 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Res.*, 50(D1):D1–D10, jan 2022.

[35] Marylyn D. Ritchie, Emily R. Holzinger, Ruowang Li, Sarah A. Pendergrass, and Dokyoon Kim. Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.*, 16(2):85–97, 2015.

[36] Matthew E. Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, 43(7):e47–e47, apr 2015.

[37] Nikola Simidjievski, Cristian Bodnar, Ifrah Tariq, Paul Scherer, Helena Andres Terre, Zohreh Shams, Mateja Jamnik, and Pietro Liò. Variational Autoencoders for Cancer Data Integration: Design Principles and Computational Practice. *Front. Genet.*, 10:1205, dec 2019.

[38] Lars Ståhle and Svante Wold. Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study. *J. Chemom.*, 1(3):185–196, jul 1987.

[39] Damian Szklarczyk, Annika L Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, Nadezhda T Doncheva, John H Morris, Peer

Bork, Lars J Jensen, and Christian von Mering. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, 47(D1):D607–D613, jan 2019.

[40] Pawel Tacik, Sadie Curry, Shinsuke Fujioka, Audrey Strongosky, Ryan J. Uitti, Jay A. van Gerpen, Nancy N. Diehl, Michael G. Heckman, and Zbigniew K. Wszolek. Cancer in Parkinson's disease. *Parkinsonism Relat. Disord.*, 31:28–33, oct 2016.

[41] Hélio A G Teive, André R Troiano, Francisco M B Germiniani, and Lineu C Werneck. Flunarizine and cinnarizine-induced parkinsonism: a historical and clinical analysis. *Parkinsonism Relat. Disord.*, 10(4):243–5, jun 2004.

[42] Johan Trygg and Svante Wold. O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. *J. Chemom.*, 17(1):53–64, 2003.

[43] David S. Wishart, Craig Knox, An Chi Guo, Savita Shrivastava, Murtaza Hassanali, Paul Stothard, Zhan Chang, and Jennifer Woolsey. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, 34(Database issue):D668–D672, jan 2006.

[44] Wei-lan Yin, Wei-guo Yin, Bai-sheng Huang, and Li-xiang Wu. Neuroprotective effects of lentivirus-mediated cystathionine-beta-synthase overexpression against 6-OHDA-induced parkinson's disease rats. *Neurosci. Lett.*, 657:45–52, sep 2017.