Article

Common Factors in Neurodegeneration:

A Meta-Study revealing Shared Patterns on a Multi-Omics Scale

Nicolas Ruffini 1§, Susanne Klingenberg1§ Susann Schweiger1 and Susanne Gerber 1,*

1 Institute for Human Genetics, University Medical Center, Johannes Gutenberg University, Mainz, Germany
* Correspondence: sugerber@uni-mainz.de;
§ Both authors contributed equally

Abstract: Neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, Huntington's disease, and Amyotrophic Lateral Sclerosis are heterogeneous, progressive diseases with frequently overlapping symptoms characterized by a loss of neurons. Studies suggested relations between neurodegenerative diseases for many years, e.g., regarding the aggregation of toxic proteins or triggering endogenous cell death pathways. Within this study, publicly available genomic, transcriptomic and proteomic data were gathered from 188 studies and more than one million patients to detect shared genetic patterns between the neurodegenerative diseases and the analyzed omics-layers within conditions. The results show a remarkably high number of shared genes between the transcriptomic and proteomic levels for all diseases while showing a significant relation between genomic and proteomic data only in some cases. A set of 139 genes was found to be differentially expressed in several transcriptomic experiments of all four diseases. These 139 genes showed overrepresented GO-Terms and pathways mainly involved in stress response, cell development, cell adhesion, and the cytoskeleton. Furthermore, the overlap of two and three omics-layers per disease were used to search for overrepresented pathways and GO-Terms. Taken together, we could confirm the existence of many relations between Alzheimer's disease, Parkinson's disease, Huntington's disease, and Amyotrophic Lateral Sclerosis on the transcriptomic and proteomic level by analyzing the pathways and GO-Terms arising in these intersections. The significance of the connection between the transcriptomic and proteomic data for all four analyzed neurodegenerative diseases showed that exploring these omics-layers simultaneously holds new insights that do not emerge from analyzing these omics-layers separately. Our data therefore suggests addressing human patients with neurodegenerative diseases as complex biological systems by integrating multiple underlying data sources.

Keywords: Multi-Omics, Alzheimer's Disease, Parkinson's disease, Huntington's disease, Amyotrophic Lateral Sclerosis, Neurodegeneration)

1. Introduction

Neurodegenerative diseases (NDD), including Alzheimer's disease, Parkinson's disease, Amyotrophic Lateral Sclerosis, and Huntington's disease, are heterogeneous, progressive
diseases characterized by a loss of neurons, accumulation of aggregated and misfolded proteins (Rubinsztein, 2006; Serrano-Pozo et al., 2011; Hussain et al., 2018; Esteves and Cardoso, 2020), cognitive decline and locomotive dysfunction (Xie et al., 2014; Dugger and Dickson, 2017; Gan et al., 2018). Despite decades of research and considerable progress in identifying risk genes, potent biomarkers, and environmental risk factors, this progression cannot be impeded. As details regarding the various (patho)physiological processes associated with Neurodegenerative Diseases (NDD) remain unclear, the diseases are still incurable.

It became generally accepted that the underlying mechanisms are polyfactorial and depend on the complex interplay of multiple (partly unknown) genetic and non-genetic variables (Mahalingam and Levy, 2014; Hinz and Geschwind, 2017; Ghasemi and Brown, 2018; Reed et al., 2019; Bellenguez, Grenier-Boley and Lambert, 2020). Excessive immune response and inflammation (Kim, Seo and Suh, 2004; Heneka, Kummer and Latz, 2014; Doty, Guillot-Sestier and Town, 2015; Heckmann, Tummers and Green, 2019), misguided apoptosis and autophagy (Bredesen, Rao, and Mehlen, 2006; Guo et al., 2018), dysfunction in mitochondria (Silva et al., 2017; Briston and Hicks, 2018; Esteves et al., 2018) and ion channels (Li and Lester, 2001; Kumar et al., 2016) and various forms of cell stress have been recognized to play a major role in neurodegenerative processes (Jomova et al., 2010; Ramanan and Saykin, 2013; Cabral-Miranda and Hetz, 2018).

By far the most prevalent of these diseases, Alzheimer's disease (AD), is an inexorably progressive brain disorder that affects higher cognitive functions. Memory loss is the typical sign of the disease. Still, there is also a significant decline in other cognition domains like language, visual-spatial skills, practical skills, reasoning, and judgment capability. (Ballard et al., 2011; Scheltens et al., 2016; Joe and Ringman, 2019). The accumulation of abnormally folded extracellular β-amyloid (senile plaques) and intracellular phosphorylated tau (neurofibrillary tangles) proteins are the distinctive pathological hallmarks of the disease (Braak and Braak, 1991; Mandelkow and Mandelkow, 1998; Iqbal and Grundke-Iqbal, 2002). Plaques and tangles have been shown to interfere with calcium signaling and synaptic transmission (Li and Lester, 2001; Sheng, Sabatini and Südhof, 2012; Berridge, 2014; Perdigão et al., 2020), to induce a persistent inflammatory response, and to lead to synapse loss and ultimately neuronal degeneration (Hashimoto et al., 2003; Wyss-Coray, 2006; Kinney et al., 2018). Inflammation caused by a central nervous system infection could be an additive factor in the pathogenesis (Itzhaki et al., 2004; Maheshwari and Eslick, 2015). Changes primarily occur in the entorhinal cortex and hippocampus, and then spread in the frontal cortex, amygdala, basal forebrain, and brainstem (Rademakers and Rovelet-Lecrux, 2009; Halliday, 2017).

Parkinson's disease (PD) is the second most common neurodegenerative disorder, mainly affecting the motor system (Dauer and Przedborski, 2003; Wright Willis et al., 2010). The aggregation of α-synuclein into Lewy bodies and Lewy neurites, mainly in the substantia nigra pars compacta and the resulting loss of dopaminergic neurons leads to distinctive symptoms including resting tremors, bradykinesia, stooped posture, and in some cases, dementia (Berardelli et al., 2001; Alexander, 2004; Jankovic, 2008). This classical view of disease origin has now been broadened to other non-motor related brain regions. There is evidence that areas like the autonomic and enteric nervous system (Knudsen and Borghammer, 2018), the olfactory bulb (Niu et al., 2018), the medulla oblongata...
(Pyatigorskaya et al., 2016), and pontine tegmentum may be affected by neurodegeneration even before the involvement of the substantia nigra. In contrast, the neocortex might be involved later throughout the disease (Novellino et al., 2020). This observation is supported by some non-motor symptoms observed in patients with Morbus Parkinson, including olfactory loss, sleep disturbance, depression, autonomic dysfunction, and cognitive impairment (Braak et al., 2003; Cerasa, Novellino, and Quattrone, 2016).

Huntington’s disease (HD) is a progressive neurodegenerative disease that manifests pathologically with the significant loss of the striatum’s GABAergic medium-sized spiny neurons (Ross and Tabrizi, 2011; Brandt, 2018). Symptoms appear in midlife and include chorea, cognitive decline, psychiatric disorders, and depression. The disease is fatal within 15 - 20 years after onset (Damiano et al., 2010; McQuade et al., 2014). While both, familial and sporadic forms of AD and PD exist, HD is an autosomal dominant neurodegenerative disease caused by the expansion of a CAG repeat in exon 1 of the huntingtin gene translating into a polyglutamine (polyQ) expansion in the N-terminus of the Huntingtin protein (Walker, 2007; Johnson and Davidson, 2010; Roos, 2010). The length of the repeat negatively correlates with the age of onset. More than 40 CAG repeats being fully penetrant with adult-onset while more than 70 repetitions generally result in juvenile-onset (Roos, 2010; Ross and Tabrizi, 2011; Pringsheim et al., 2012). Physiologically, huntingtin seems to have multiple biological functions, including axonal and vesicular transport, endocytosis, post-synaptic signaling, and cell survival pathways (Harjes and Wanker, 2003; Cattaneo, Zuccato and Tartari, 2005). The mutant huntingtin gene is prone to cleavage. The resulting short fragments containing the N-terminal polyglutamine expansion oligomerize and form aggregates that have been implicated in neurotoxicity (Wanker et al., 2019). Furthermore, dysregulation of several functions occur, including gene transcription (Cha, 2000; Roze et al., 2008), axonal transport of critical factors (Gunawardena et al., 2003), calcium and potassium signaling (Mackay, Nassrallah and Raymond, 2018; Zhang, Wan and Tong, 2018), protein-protein interactions, autophagy (Ochaba et al., 2014; Martin et al., 2015), and proteasomal and mitochondrial capacities (Damiano et al., 2010; Reddy and Shirendeb, 2012; McQuade et al., 2014). Even though the huntingtin gene is expressed ubiquitously, the striatum and cortical areas are the most affected regions as shown in postmortem histological evaluation (Ferrante et al., 1987; Tang et al., 2007; Ross and Tabrizi, 2011) or non-invasive brain magnetic resonance imaging (Aylward et al., 2004; Kassubek et al., 2004). The severity of striatal alterations is correlated with the degree of motor, cognitive, and psychiatric perturbations, suggesting that striatal degeneration is an essential aspect of HD physiopathology (Myers et al., 1988). Pre-symptomatic patients (i.e., carrying the mutation but are still asymptomatic) already show significant atrophy of the caudate and putamen (Aylward et al., 2004; Kassubek et al., 2004). This suggests that degenerative events like cell shrinkage or loss begin years before the occurrence of clinical symptoms.

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease affecting both the upper and lower motor neurons (Mitchell and Borasio, 2007; Wijesekera and Leigh, 2009; Kiernan et al., 2011). It is characterized by progressive muscular paralysis reflecting the degeneration of motor neurons in the primary motor cortex, corticospinal tracts, brainstem, and spinal cord (Kang et al., 2013; Philips and Rothstein, 2014; Novellino et al., 2020). The mean age of onset for sporadic ALS is about 60 years. Approximately two-thirds of patients with typical ALS have a spinal form of the disease (limb onset) where muscle weakness starts either distally or proximally in the upper and lower limbs (Turner et al., 2011;
Robberecht and Philips, 2013). Patients with bulbar onset ALS usually suffer from dysarthria and dysphagia, and limb symptoms can develop almost simultaneously with bulbar symptoms (Kiernan et al., 2011; Cistaro et al., 2012; Guareschi et al., 2012). Paralysis is progressive and leads to death due to respiratory failure within 2–5 years (Mitchell and Borasio, 2007; Kiernan et al., 2011). Most ALS cases are sporadic, but 5–10% of cases are familial with mutations of the \textit{SOD1} gene and the \textit{TARDBP} gene (TDP-43) (Wijesekera and Leigh, 2009; Renton, Chiò and Traynor, 2014). The antioxidant enzyme SOD1 protects cells from the damaging effects of superoxide radicals. Deleterious mutations of \textit{SOD1} lead to the accumulation of highly toxic hydroxyl radicals, which causes degradation of both nuclear and mitochondrial DNA and protein misfolding (Guareschi et al., 2012; Kaur, McKeown and Rashid, 2016). The transactivating response DNA binding protein of 43 kD (TDP-43) is a highly conserved and ubiquitously expressed protein localized primarily within the nucleus. It is important in gene transcription and RNA splicing regulation (Lagier-Tourenne, Polymenidou, and Cleveland, 2010; Prasad et al., 2019), but also regulates the activity of retrotransposons (Savage et al., 2018; Tam et al., 2019). It is predominantly nuclear, but shuttles in and out of the cytoplasm and along axons, where it can aggregate to hyperphosphorylated and ubiquitinated TDP-43 deposits both in the familial and the sporadic form of ALS (Prasad et al., 2019).

Overlap of phenotypic traits of the NDDs described above suggests common pathogenic mechanisms underlying distinct NDD. Compared to studying individual diseases separately, identifying and analyzing the common dysfunctional proteins and dysregulated pathways of the diseases might elucidate the fundamental insights into their pathogenic process (Li, Nie and Yu, 2015). It was previously shown that there is no overlap between AD, PD, and ALS on genomic data and some shared pathways for AD, PD, ALS, and HD transcriptomic data (Arneson et al., 2018), but proteomic data and the latest entries in the databases were not considered. Besides looking for overlapping genes between the different NDD or omics layers, we also analyzed whether this number is sufficiently high to claim a significant relationship between NDD or omics layers. An overview of the methodologic procedure is given in Figure 2. By investigating 188 studies (see Figure 1) in total, this study can detect a stable signal that arises in mainly late-stage NDD across tissues, methods, and omics-layers and, therefore, can help unravel underlying patterns across neurodegenerative diseases.
Figure 1 Overview of the analyzed NDD and omics-layers. Data from 188 studies, including more than 1 million human samples, were gathered to find common patterns of genes, pathways, and GO-terms connecting AD, PD, ALS, and HD. The area of each disease's fraction in the pie chart is just visualization and does not represent the correct number of studies per disease and omics-layer. For the absolute numbers per disease and omics-layer see and Table 2.
Figure 2 Workflow Overview: Data Acquisition was performed using the GWAS-catalog for genomic data, the EMBL-EBI Expression Atlas, the GEO database for transcriptomic data, and manual PubMed and Google Scholar research for proteomic data. After filtering these raw data tables and applying some data transformation, the processed data were used for the data analysis. For every omics-layer in each disease, Gene occurrence bar plots were created for the 20 most abundant genes. For detecting overlapping gene names on multiple omics-layers, stacked bar plots were created for every disease, showing the average direction of regulation for the transcriptomic and proteomic layer. The intersections were visualized as Venn diagrams of all combinations on the disease and the omics-layer level. Finally, each set of genes after the intersections was used for KEGG pathway and GO-Term analyses.
Results

**Gene Occurrence:** The 20 most abundant genes per disease for the transcriptomic and proteomic data are shown in colored bar plots representing the abundance and mean direction of regulation (see Equation 1) of each gene (see Figure 3). The overall most abundant gene was GFAP in the proteomic AD data. Most of the twenty most abundant AD genes appeared more than ten times, whereas none of HD and ALS genes did appear at least ten times. PD data showed some highly abundant genes in the transcriptomic data – ANK1 appeared ten times, RPL21 appeared 14 times. Interestingly, none of the twenty most abundant genes in the PD transcriptomic data showed a positive mean direction of regulation. The most abundant genes for the other diseases and omics-layers showed more heterogeneous mean directions of regulation.

**Intersection:** To quantify, if the number of overlapping genes is high between omics-layers for one disease and between diseases for one omics-layer, a hypergeometric test was performed. Analyzing the number of overlapping genes between transcriptomic and proteomic data showed a significant number of overlapping genes for all four diseases (see Figure 4). The number of overlapping genes between the genomic level and the other two omics-layers was not significantly enriched for ALS and HD, and no single gene occurred in the intersection between the genomic and the proteomic level for these diseases. For AD, all pairwise intersections between these three omics-layers showed a significantly enriched number of overlapping genes. For PD, the number of overlapping genes between genomic and transcriptomic data was significantly enriched, but no significant overlap could be seen for the proteomic – genomic overlap. AD showed the most considerable overlap between all three omics-layers with 67 genes. For PD, three genes were found in the intersection between all three analyzed omics-layers. The number of genes found in the GWAS catalog was the highest for AD with 434 single nucleotide polymorphisms (SNPs). For PD, 218 SNPs were found; 68 for ALS and 34 SNPs for HD. The number of overlapping SNPs between each pair of diseases ranged from zero to eleven and was significantly high for the pairwise overlaps between AD and PD as well as AD and ALS in a hypergeometric test (see Figure 5).

For the transcriptomic data, AD again showed the highest number of genes with a total of 14,737 genes that were differentially expressed in at least two experiments. PD showed 4,713 differentially expressed genes, ALS showed 897 and 4,249 differentially expressed genes could be found for HD. All pairwise comparisons of diseases on the transcriptomic level showed highly significant enrichment in the number of overlapping genes. One thousand nine hundred sixty-four gene names could be related to differentially expressed proteins for AD. For PD, we found 434 gene names, 155 for ALS, and 104 for HD. All pairwise overlaps between the four diseases were highly significantly enriched for the proteome data.
Figure 3 Gene Occurrence Bar Plot showing the occurrences of the twenty most abundant genes in terms of experiments they appeared in as significantly up- or downregulated on the transcriptomic (left column) and proteomic layer (right column) for the four analyzed NDDs (from top: AD, PD, ALS, HD). The color of each bar gives the mean direction of regulation.
Figure 4 Venn Diagrams and tables are showing intersections and the hypergeometric test results for the overlap between transcriptome and proteome data (column 1), transcriptome and SNP data (column 2), and proteome and SNP data (column 3) for AD, PD, ALS, and HD. All overlaps between the transcriptomic and proteomic levels were significantly higher than expected. AD also showed a significant overlap between the genomic and transcriptomic/proteomic level, PD for genomic-transcriptomic. The other intersections with genomic data did not show significantly more genes than expected by chance.
Figure 5 Venn diagrams and hypergeometric test results for the overlap between significantly differentially expressed genes on the genomic level (top), transcriptomic level (middle), and proteomic level (bottom) for AD, PD, ALS, and HD. All tested intersections show highly significant enriched numbers of overlapping genes for the transcriptomic and proteomic data. The genomic data showed significantly enriched numbers of overlapping genes for the AD-PD and AD-ALS intersections.
GO-Term- & Pathway-Analyses: As the number of possible combinations and analyses was very high this study concentrated on describing the analyses of the genes appearing in the intersection of all diseases per omics layer and on those, appearing in the transcriptomic and proteomic data per disease (+ genomic data for AD). GO-Terms and genes were classified according to the conceptual model of candidate pathways contributing to neurodegeneration shown in Figure 6 (modified version according to (Ramanan and Saykin, 2013)). NDD can be classified into broader functional groups based on their primary site or mode of action (intracellular mechanisms, local tissue environment influences, and systemic influences). Candidate pathways influencing the balance of neuronal survival, and degeneration are shown below. This model’s pathways and mechanisms are highly related and can have overlapping or interacting components that can collectively modulate neurodegenerative processes.

![Figure 6](image)

**Figure 6** Classification of candidate pathways contributing to neurodegeneration into three groups according to their cellular mechanisms or their primary site of action (modified and extended version) (Ramanan et al., 2014). All of our found GO-Terms Biological Processes (BP) were involved in at least one of the given categories’ intracellular mechanisms, local tissue environment, or systemic environment.

The intersection of genomic, transcriptomic, and proteomic data

AD: The over-representation analysis (ORA) of the gene set emerging from the interaction of the genomic, transcriptomic, and proteomic AD data showed 18 GO-Terms that are significantly overrepresented (FDR ≤0.05) in the category Biological Process (BP). By affinity propagation, they are categorized into altered autophagy and immune response (ß-amyloid formation), stress response/neuronal death, and extracellular matrix/cell communication. All eight Cellular Components (CC)-Terms are related to extracellular matrix and cell migration. Both Molecular Function (MF)-Terms belong to metabolism (lipoprotein binding).
Intersection of transcriptomic and proteomic data

**AD:** There are many more genes overlapping if only transcriptomic and proteomic data are intersected, and thus much more BP-Terms (about 1350) can be found. By affinity propagation, BP was reduced to about 400 terms. The majority of these terms deal with cell adhesion and cell differentiation and metabolism, especially sugar- and lipid metabolism. Signal transduction (synapsis-related processes) was also frequently overrepresented, followed by stress response, autophagy and apoptosis, mitochondrial functions, signal transduction (ion channels), and immune system. The circa 270 overrepresented CC-Terms (75 after affinity propagation) were consequentially high in endocytosis (vesicles and membranes), signal transduction (neurons and synapses), mitochondrial functions, and other organelles and cell adhesion. More than half of the 48 MF-Terms deal with metabolism (protein binding and lipid transport), whereas the remaining terms are distributed between cell adhesion, signal transduction, and stress and immune response.

**PD:** The approx. 200 overrepresented BP-Terms found for the intersection of transcriptomic and proteomic PD data were classified by affinity propagation to cytoskeleton and cell development (10/36), metabolism (9/36), almost half of it concerning sugar metabolism, stress, and immune response (6/36) and transcription regulation (4/36). The rest dealt with apoptosis, signal transduction (synapses, ion channels), and organs' development (muscle cells and neurons of the substantia nigra). The 75 significantly overrepresented CC-Terms were reduced to 11 categories by affinity propagation, i.e., endocytosis (vesicles, especially in synapses) (5/11), mitochondrial functions (3/11), and cell adhesion (3/11). From 49 molecular functions, 11 terms with very diverse processes like autophagy (protein tagging for proteasome) and cytoskeleton (especially tubulin-binding) and immune system, translation regulation, and metabolism emerge affinity propagation.

**ALS:** Most of the overrepresented BP-Terms deal with cell adhesion and cytoskeleton (7/8). Only one process is immune response (Staphylococcus aureus infection). All five overrepresented CC-Terms are connected to cell differentiation (focal adhesion, extracellular matrix), and all 3 MF-Terms refer to the cytoskeleton.

**HD:** The 21 overrepresented GO-Terms for BP can be classified by affinity propagation into the categories: cell development, signal transduction, and metabolism, especially lipid transport; all three CC-Terms are associated with cell adhesion, and both MF-Terms concern metabolism, especially lipid transport.

**Intersection of AD, PD, ALS, and HD**

**Transcriptome:** The 28 overrepresented BP-Terms for the intersection of all four diseases' transcriptomic data can be classified by affinity propagation, mainly to stress response (5/8). The remaining processes are concerning transcriptional regulation and extracellular matrix. All of the six detected CC-Terms are related to focal adhesion. The six overrepresented MF-Terms are mainly about cell adhesion and cytoskeleton (4/6), vascular factors (growth factor binding) (2/6), and stress response (1/6).

**Proteome:** For the intersection of proteomic data for all four analyzed NDD, three significantly overrepresented GO-Terms were found. They are all relevant to metabolism (maintenance of protein stability) and the immune system (humoral immune response).
4. Discussion

**AD – Transcriptome:** Among the 20 top hits we found for the differentially expressed genes in AD are ELF1, MSI2, and TBL1X, which regulate transcription, BAZ1A for chromatin remodeling, DNAJC1 as a response to heat stress, PAK1, and ANTRX1 involved in cell adhesion and cell migration and GDAP1L1 for neuronal development. The most intriguing finding is the transcript of the nicotinamide nucleotide adenylyltransferase 2 (NMNAT2) being downregulated in 16 studies, according to the literature (Ali et al., 2016; Bennett and Keeney, 2018). NMNAT2 is an essential enzyme for energy metabolism in the brain, catalyzing an essential step in NAD/NADP biosynthesis. Its expression levels correlate positively with cognitive function and negatively with AD pathology. NMNAT2 appears to be required for axon survival and functions as a chaperone to aid in the refolding of misfolded proteins to reduce proteotoxic stress (Zhai et al., 2008; Lukacs et al., 2019).

**AD - Proteome:** The top 20 hits for the genes related to the differentially expressed proteins in AD are mainly involved in cytoskeleton organization, in cell adhesion and neuronal activity. The microtubule-associated protein tau (MAPT) is expressed at its highest levels in neurons throughout the central nervous system. It is involved in assembling and stabilizing the cytoskeleton (Strang, Golde and Giasson, 2019). Microtubules help cells maintain their shape, assist in cell division, and are essential for transporting materials within cells. One of tau’s primary functions is to bind to and promote microtubules' assembly and stability; this binding activity can be negatively regulated by phosphorylation at select sites (Arendt, Stieler and Holzer, 2016). An imbalanced regulation in protein kinases and protein phosphatases is the direct cause of AD-like tau hyperphosphorylation (Wang and Liu, 2008) and hyperphosphorylated tau accumulation induces synaptic toxicity and cognitive impairments (Ma et al., 2017). Clusterin (CLU) or APOJ is a multifunctional glycoprotein that has been implicated in several biological processes, including lipid transport, membrane recycling, cell adhesion, programmed cell death, and the complement cascade. Although Clusterin usually is a secreted protein, it has also been found intracellularly under certain stress conditions (Koltai, 2014). Clusterin's ability to interact and bind to Aβ appears to alter aggregation and promote Aβ clearance, suggesting a neuroprotective role (DeMattos et al., 2004). CLU is considered the third most significant genetic risk factor for the non genetic form of AD, after APOE and BIN1 with reduced secreted Clusterin as a mechanism for Alzheimer-associated CLU mutations (Bettens et al., 2015). Interestingly, in most of our studies, CLU was upregulated.

**PD - Transcriptome:** Most of the differentially expressed genes found in PD were involved in gene expression regulation and stress response/autophagy processes. Among the top 20 differentially expressed genes in PD, downregulated ANK1 was found. Ankyrins are a family of proteins that link the integral membrane proteins to the underlying spectrin-actin cytoskeleton and play critical roles in cell motility, activation, proliferation, contact, and specialized membrane maintenance domains. They attach integral membrane proteins to cytoskeletal elements. ANK1 shows altered methylation and expression in Alzheimer’s disease (De Jager et al., 2014; Lunnon et al., 2014) and Parkinson’s disease (Liscovitch and French, 2014). Another interesting finding is the enzyme pyridoxal kinase (PDXK) that catalyzes the conversion of vitamin B6 (pyridoxine, pyridoxal, and pyridoxamine) to pyridoxal 5’-phosphate (PLP), which acts as a cofactor for over 140 different enzymatic
reactions. Of direct relevance to PD, the second step in the biosynthesis of dopamine by the enzyme dopa decarboxylase is dependent on PLP as a cofactor and becomes rate-limiting in patients receiving L-DOPA therapy (Tan et al., 2005). An adaptive mechanism may explain the upregulation of PDXK in dopaminergic neurons to increased dopamine metabolism in the remaining functional dopaminergic neurons of the SN or L-DOPA therapy (Elstner et al., 2009). Commonly, however, as in our studies, PDXK is downregulated in PD patients.

**PD - Proteome:** Like the differentially expressed genes most of the differentially expressed proteins were involved in gene expression regulation and stress response/autophagy processes but also in tubulin polymerization. Microtubules are essential for the mitotic spindle, the extracellular matrix, cell migration, and the growth of the myelin sheath protecting neurons.

One of the enzymatic reactions where the above mentioned PLP acts as a cofactor is the catabolism of gamma-aminobutyric acid (GABA) into succinic semialdehyde. In this reaction, PLP is bound to the 4-aminobutyrate aminotransferase (ABAT), which belongs to the top 20 genes related to the differentially expressed proteins found in our study. GABA is an important, mostly inhibitory neurotransmitter in the central nervous system, estimated to be present in nearly one-third of human synapses. Parkinson patients with ABAT deficiency showed increased GABA levels in the plasma and cerebrospinal fluid and associated symptoms such as encephalopathy, psychomotor retardation, seizures, hypotonia, hyperreflexia, lethargy, and abnormal electroencephalogram (Kim et al., 2018; O’Gorman Tuura, Baumann and Baumann-Vogel, 2018). Another interesting differentially expressed protein in PD is the tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein epsilon YWHAE. This protein belongs to the 14-3-3 family, it is especially abundant in brain tissue (Boston, Jackson and Thompson, 1982), comprises several isoforms and interacts mainly with phosphorylated protein partners. Members of the protein family have been associated with the development of neurodegenerative diseases (Berg, Holzmann and Riess, 2003), among them Parkinson’s disease and Alzheimer’s disease (Umahara et al., 2004), where they are found in neurofibrillary tangles (NFTs). NFTs are composed of hyperphosphorylated Tau protein displaying paired helical filaments and substantial amounts of 14-3-3 proteins. It could be shown that the functional and direct interaction of overexpressed 14-3-3 with Tau proteins leads to Tauopathies, such as Tau accumulation and Tau detachment from microtubules (Joo et al., 2015).

**HD – Transcriptome:** The top 20 differentially expressed genes in HD are mainly involved in the mitotic cell cycle, epigenetic modifications, cell differentiation and proliferation, apoptosis, and immune responses to inflammation and infection. In most of our studies, Perilipin 2 (PLIN2) was found to be upregulated. PLIN2 is a perilipin family member, which coat intracellular lipid storage droplets and develops and maintains adipose tissue. However, it is not restricted to adipocytes but is also found in astrocytes in the brain. The lipid imbalance plays a significant role in neurodegeneration, such as in Alzheimer’s disease (Dalhaimer, 2019). The E4 allele of APOE, gene product of which is a major structural component of low-density lipoproteins, is a decisive genetic risk factor for developing late-onset Alzheimer’s disease (Liu et al., 2013). Astrocytes expressing E4 accumulate significantly more and smaller lipid droplets compared to E3 astrocytes. Accordingly, the expression of perilipin-2 was higher in E4 astrocytes (Farmer, Kluemper, and Johnson, 2019). Another exciting gene found in HD is Kinectin 1 (KTN1), also known
to be related to PD. KTN1 encodes an integral membrane protein, primarily localized in the endoplasmic reticulum membrane. This protein binds kinesin, a motor protein that is involved in intracellular organelle motility. Powered by the hydrolysis of ATP, kinesins move along microtubule filaments. The active movement of kinesins supports several cellular functions, including mitosis, meiosis, and cellular transport, such as in axonal transport. Most kinesins move towards the plus end of a microtubule, transporting cargo such as protein and membrane components from the inner cell towards the periphery. KTN1 facilitates vesicle binding to kinesin, regulating crucial developmental processes including axonal guidance, vesicular transport of molecules, and apoptosis (Kumar, Yu, and Sheetz, 1995; Hibar et al., 2015), as well as neuronal cell shape and neuronal migration through kinectin–kinesin interactions (Zhang et al., 2010). Neurons with more kinectin1 have larger cell bodies (Toyoshima and Sheetz, 1996). A positive relationship among PD risk and KTN1 mRNA expression in the putamen and putamen volumes, and a modest relation between PD risk and KTN1 mRNA expression in SNc could be shown, suggesting that KTN1 may play a functional role in the development of PD (Mao et al., 2020).

**HD – Proteome:** The genes related to the differentially expressed proteins could be linked to mitochondrial and lysosomal functions, and the actin-, dystrophin- and kinesin filaments of the cytoskeleton and muscle cells. Among the proteins that were upregulated in HD, we found the Cysteine and Glycine Rich Protein 1 (CSRPI), which may be involved in regulatory processes essential for development and cellular differentiation. CSRPI’s LIM/double zinc-finger motif occurs in proteins with critical functions in gene regulation, cell growth, and somatic differentiation. Members of the CRP family are involved in a late stage in muscle differentiation (Pomès, Louis and Beckerle, 1997), and it could play a role in neuronal development (Knierim et al., 2016). Laminin Subunit Gamma 1 (LAMC1), a high-molecular-weight glycoprotein of the extracellular matrix, which belongs to the family of laminins, was found to be downregulated in all our HD studies. Laminins are the principal non-collagenous constituent of the basal lamina, a protein network foundation for most cells and organs, influencing cell differentiation, migration, and adhesion (Durbeej, 2010). They are also an integral part of the structural scaffolding in almost every organism’s tissue and are incorporated and secreted into cell-associated extracellular matrices. Defective laminins can cause improperly formed muscles, thus leading to a form of muscular dystrophy (Yurchenco and Patton, 2009).

**ALS – Transcriptome:** The main pathways for the top 20 genes of the differentially expressed genes for ALS are cell cycle and gene expression regulation, stress response, apoptosis, the maintenance of the cytoskeleton, and the extracellular matrix neuronal development. CD44, also present among the top 20 differentially expressed HD genes, was found to be upregulated in ALS, too. Its protein is a transmembrane, highly glycosylated protein involved in cell-cell interactions, cell adhesion, and migration, helping cells sense and respond to tissue microenvironment (Dzwonek and Wilczyński, 2015). In the nervous system, CD44 expression occurs in both neuronal and glial cells (McKenzie, Dalchau, and Fabre, 1982). The role of CD44 in the physiology and pathology of the nervous system is not entirely understood. However, there exists evidence suggesting it might be involved in the axon guidance, cytoplasmic Ca²⁺ clearance, dendritic arborization, synaptic transmission, epileptogenesis, oligodendrocyte and astrocyte differentiation, and post-traumatic brain repair (Jones et al., 2000; Matzke et al., 2007; Vedunova et al., 2013; Skupien et al., 2014). Aquaporin 1 is another crucial ubiquitous protein that was among the top 20
differentially expressed genes for ALS. Aquaporins are a family of widely distributed membrane-inserted water channel proteins providing a pathway for osmotically driven water, glycerol, urea, or ions transport through cell membranes and thus control particular aspects of homeostasis. Aquaporins, especially AQP4, are abnormally expressed in the Central Nervous System under some pathological conditions like neurodegenerative diseases in which preservation of brain homeostasis is at risk (Foglio and Luigi Fabrizio, 2010). AQP1 encodes a small integral membrane protein, which is mostly found in the plasma membranes of red cells and kidney proximal tubules with high permeability to water. Herewith, water can be hold even against the direction of an osmotic gradient (Badaut et al., 2014).

**ALS – Proteome:** Among the genes related to ALS’s differentially expressed proteins were secreted phosphoprotein 1 (SPP1) or osteopontin and hemopexin (HPX) genes. SPP1 is a matrix phosphoprotein expressed by various tissues and cells, including the immune system and the nervous system. Immune cells, such as macrophages and T lymphocytes, are essential sources of SPP1 during inflammatory processes (Mousavi, Agah, and Tafakhori, 2020). SPP1 may have a role in neurodegenerative diseases, including multiple sclerosis, PD, AD, and frontotemporal dementia (Carecchio and Comi, 2011; Agah et al., 2018). It was reported that SPP1 was selectively expressed in alpha motor neurons, which are the most vulnerable in ALS (Misawa et al., 2012; Yamamoto et al., 2017). Therefore, it may also play a role in the pathogenesis of ALS. Hemopexin encodes a plasma glycoprotein that binds heme with high affinity. It is an acute-phase protein that transports heme from the plasma to the liver for breakdown and iron recovery. It may be involved in protecting cells from oxidative stress (Ashraf et al., 2020), because Hb oxidation liberates free heme, which is a source of redox-active iron-producing reactive oxygen species, inducing lipid peroxidation (Papanikolaou and Pantopoulos, 2005). This toxicity is modulated by hemopexin. Proteomics studies demonstrated increased plasma and cerebrospinal fluid hemopexin in AD patients compared to cognitively normal controls, suggesting impaired compensation in neurodegeneration (Castano et al., 2006). It appears that surplus contribution to heme by the breakdown of Hb may overwhelm the capacity of the heme scavenging system in AD (Smith and McCulloh, 2015; Ma et al., 2016).

The analysis of the twenty most abundant genes in each NDD’s transcriptome and proteome data interestingly showed many highly abundant genes in single NDD that were also described for other NDD, e.g., SPP1, that was highly abundant in the ALS proteomic data but already described for PD, AD and other NDD like multiple sclerosis. Taken together, many of the upcoming genes were related to cytoskeleton, cell adhesion, extracellular matrix formation, apoptosis, and stress response. The results show the general conformity of the most abundant genes with the current study situation. Many genes are described as playing essential roles in more than one of the investigated NDD. Additionally, looking at the mean direction of regulation among all studies involving a particular gene on the respective omics-layer showed that in some cases, the current study situation seems to be contradictory even for some of the most abundant significantly differentially expressed genes, such as ANTXR1 and ARHGAPI2 for transcriptomic AD data and ZFAND5 and PCMT1 for PD transcriptomic and proteomic data. However, apart from these contradictory regulated genes, most of the twenty top abundant genes were consistent among most studies.
**Intra-disease relations:** The intersection of genes found in genomic, transcriptomic, and proteomic data per disease showed more surprising results. HD and ALS did not show any genes that were shared between the genomic and proteomic data. The number of five genes shared between these two omics-layers for PD is not higher than expected by chance. Only in AD a significant overrepresentation of genes shared between the genomic and proteomic layer was found. On the other hand, both AD and PD showed a significantly high number of genes in the genomic and transcriptomic data (p-value\textsubscript{PD}: 5.38E\textsuperscript{-4}). The number of shared genes between the transcriptomic and proteomic level was highly significantly enriched in all four investigated NDD.

**Inter-disease relations:** Interestingly, the number of shared genes between the four NDD on the genomic level was significantly enriched only for the AD-PD and AD-ALS comparisons. For the proteomic and transcriptomic data, all numbers of pairwise overlapping genes were significantly enriched. However, the overlap between all four analyzed NDD was large enough to be further analyzed only at the transcriptomic level. The genomic and proteomic data of all four NDD showed no (genomic) and four (proteomic) overlapping genes. Nevertheless, the overlap between all four NDD on the transcriptomic level showed 139 genes. The overrepresented GO-Terms for biological process in the genes appearing in the intersection of all four NDD on the transcriptomic level showed results that are partially concordant to a meta-study of AD, PD and ALS from 2019 (Bayraktar et al., 2019), that analyzed raw data of 259 individuals. They found pathways and biological processes associated to heat shock proteins, cellular response to heat, stress response, GABA synthesis and Protein folding as overrepresented in the four used datasets and stated the importance of heat shock proteins (HSP) as a general target of NDD (Kampinga, Harm H., 2016) and the importance of HSP associated pathways in HD (Kaltenbach et al., 2007).

Our GO-Term analysis revealed biological processes that are associated to cellular response to heat, stress response and protein folding. However, no significant overrepresentation was found for HSP or GABA synthesis in our analysis of AD, PD, ALS and HD. We furthermore found biological pathways associated to response to hypoxia and cell ageing as overrepresented in the gene set that was common to all four NDD on the transcriptomic level. Consequently, we could confirm the findings regarding heat response, protein folding and stress response, although our analysis did neither explicitly show pathways or biological processes associated to HSP nor GABA synthesis.

**Conclusion:**

GWA studies made a big contribution towards understanding NDD in the last 15 years, with several hundred disease-associated risk loci. However, until now for most NDD no targeted therapies have emerged from these GWA studies (Diaz-Ortiz and Chen-Plotkin, 2020). As NDD are causing vast transcriptomic changes in the ageing brain, it is crucial to take transcriptomic data analysis into account as well, when analyzing NDD (De Jager, Yang and Bennett, 2018). The further translation from transcriptomic changes to proteomic occurs only indirectly and shows just a limited correlation between mRNA and protein expression (Becker et al., 2018). Consequently, even the combination of genetic and transcriptomic data is not adequate to give a complete picture of changes taking place due to NDD. Each additional level of information can contribute to a better understanding of the complex interrelationships of these interacting omics-layers.
For addressing the challenges of such complicated diseases, the whole field of biomedicine is changing towards creating and facilitating a variety of databases and analysis pipelines for separate omics-layers as well as multi-omics integration (Manzoni, Lewis and Ferrari, 2020). Many of these pipelines are mainly data-driven and facilitate clustering and supervised machine learning techniques to find important patterns of features contributing to the identification of e.g. proteins that are associated to NDD (Yu et al., 2020) or to reveal cross-talk patterns in multi-omics data (Nguyen and Wang, 2020).

According to the necessity of approaching complex diseases with the use of multiple omics-layers and data-driven methods utilizing large amounts of data, we combined data of three omics-layers from databases and literature mining of more than one million subjects and 188 studies to show shared genes within and between the four analyzed NDD and extract the pathways and processes in which they are overrepresented.

To classify the gained information in this study, it is important to keep in mind that the transcriptomic and proteomic data were gathered from various tissues, partly different severities of diseases and using different methods. Consequently, the signals that have been found in this meta-study rather represent stable signals across tissues emerging in the late stage of NDD than subtle effects that might only be present in specific brain regions, or earlier disease states as those subtle effects would probably be canceled out when doing a GO-Term or Pathway-Analysis by the number of other signals and the fact that the majority of the used data was based on late stage NDD. The inequality between the four different diseases at the given time point concerning the study situation and the amount of insight should also be considered and was visible in the differing number of samples, studies and differentially expressed genes between AD and the other three NDD. The highly significant overrepresentation of genes in the intersection of proteomic and transcriptomic data in all investigated NDD shows the importance of simultaneously analyzing multiple omics-layers. On the other hand, despite the given overlap, many differences between the transcriptomic and proteomic layers remain, which was also shown in the differences between pathways and GO-Terms being significantly enriched in the transcriptomic layer compared to the proteomic layer. While the intersection of all NDD on the transcriptomic layer showed terms related to stress response, cell development, cell adhesion, and cytoskeleton, the genes appearing in the intersection of the proteomic data were related to protein stability and immune response. Accordingly, analyzing transcriptomic and proteomic data simultaneously in a multi-omics approach while still considering those effects exclusive to one omics-layer is necessary to investigate NDD to the full extend. Regarding future research on this topic, it might be helpful to expand the repertoire of omics-layers by epigenomics and / or adding further NDD. However, analyzing and interpreting all possible intersections was already not manageable for the current set of diseases and omics-layers with the used approaches due to the large amount of combinatorically possible intersections. Consequently, different analysis approaches might be an option to deal with the further rising degrees of freedom regarding the analysis. Concentrating further on the differences and coherences in the direction of regulation of interesting genes also holds new insight regarding the clustering of diseases or finding similarly regulated gene sets.
Materials & Methods

**Genome:** The GWAS Catalog data for Alzheimer’s disease (AD), Parkinson’s disease (PD), Amyotrophic Lateral Sclerosis (ALS), and Huntington’s disease (HD) were downloaded on the 28th of April 2020. The GWAS Catalog contains SNP data of GWAS studies for SNPs showing a statistical significance of SNP-trait p-value < 1.0E-5 in the overall population. For every SNP, data such as p-value, upstream gene(s), mapped gene, reported gene(s), and many more are stored. We focused on the genes given as "Reported Gene(s)” in the four examined diseases' full data tables for our analysis. The EFO numbers for the exact search pattern were EFO_0000249 (Alzheimer’s disease), EFO_0002508 (Parkinson’s disease), Orphanet_399 (Huntington’s disease), and EFO_0000253 (Amyotrophic Lateral Sclerosis). A table containing all studies' names and the number of investigated samples for every disease is appended in the supplement (A1: GWAS Catalog study information).

**Transcriptome:** We browsed the databases Gene Expression Omnibus (GEO) (Barrett et al., 2013) and the Expression Atlas (Papatheodorou et al., 2018). The GEO is a public data repository, in which microarray and RNA-seq datasets can be found. The keywords for the GEO database were <name of disease> AND (“microarray” OR “RNAseq”) AND “human.” The latest literature research was done in July 2020. The Expression Atlas is a service of EMBL-EBI and provides re-analyzed and manually curated data of more than 3000 experiments. It was used in release 35 (May 2020, https://www.ebi.ac.uk/gxa/home) and scanned for Alzheimer, Parkinson, Huntington, and Amyotrophic Lateral Sclerosis setting the filter to "Homo sapiens” in the section Differential Experiments. Whenever experiments with increasing severities of an NDD were conducted, we used the differentially expressed genes appearing in the comparison between the group of the most severe disease state and the control.

An overview of all used studies for gathering the transcriptomic data is given in

Table 1 Overview of the number of cases, controls, and the total number of studies per disease throughout all analyzed transcriptome studies. In total, data of 2475 samples were gathered from 46 studies analyzing transcriptomic data.

<table>
<thead>
<tr>
<th>Transcriptome</th>
<th>Case</th>
<th>Control</th>
<th>Σ of samples</th>
<th>studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>1029</td>
<td>756</td>
<td>1785</td>
<td>14</td>
</tr>
<tr>
<td>PD</td>
<td>154</td>
<td>167</td>
<td>321</td>
<td>12</td>
</tr>
<tr>
<td>ALS</td>
<td>170</td>
<td>131</td>
<td>301</td>
<td>8</td>
</tr>
<tr>
<td>HD</td>
<td>39</td>
<td>29</td>
<td>68</td>
<td>12</td>
</tr>
<tr>
<td>Σ</td>
<td>1392</td>
<td>1083</td>
<td>2475</td>
<td>46</td>
</tr>
</tbody>
</table>
We browsed publications of the last ten years in PubMed and Google Scholar with the keywords: ("neurodegenerative diseases" OR "Alzheimer* disease" OR "Parkinson* disease" OR "Huntington* disease" OR "Amyotrophic Lateral Sclerosis") AND (proteomics OR "quantitative proteomics" OR "differentially expressed proteins" OR biomarkers) AND human NOT mice. An overview of all used studies for gathering the transcriptomic data is given in Table 2. A table showing each study’s information is given in the supplement (Error! Reference source not found. Error! Reference source not found.). As well as with the transcriptomic data, we used the results emerging from the comparison of the group with the most severe disease state compared to the control, when experiments of increasing severity were available.

<table>
<thead>
<tr>
<th>Proteome</th>
<th>Case</th>
<th>Control</th>
<th>Σ of samples</th>
<th>studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>293</td>
<td>274</td>
<td>567</td>
<td>9</td>
</tr>
<tr>
<td>PD</td>
<td>302</td>
<td>279</td>
<td>581</td>
<td>7</td>
</tr>
<tr>
<td>ALS</td>
<td>471</td>
<td>691</td>
<td>1162</td>
<td>5</td>
</tr>
<tr>
<td>HD</td>
<td>360</td>
<td>181</td>
<td>541</td>
<td>3</td>
</tr>
<tr>
<td>Σ</td>
<td>1426</td>
<td>1425</td>
<td>2851</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 2 Overview of the number of cases, controls, and the total number of studies per disease throughout all analyzed proteome studies. In total, data of 2851 samples were gathered from 24 studies analyzing proteomic data.

Data Management: The raw genomic, transcriptomic, and proteomic data tables from 188 different studies were transformed into standardized tables for each disease on every omics-layer. Different conversion was applied within this data management process, such as converting fold change to log2-fold change (log2FC), log10-p-value to p-value, removal of entries with missing Gene Name or separating rows that contained several Gene Names (proteomic data). Differences in multiple testing corrections were accepted, such as differences in the exact calculation of the fold change (log2FC, G-fold Change). Only those genes with a false-discovery-ratio (FDR) ≤ 0.05 were selected after applying those conversions where necessary. Besides, all genes that appeared as differentially expressed in only one experiment on the transcriptomics or proteomics level were discarded, to further reduce the number of genes that appeared randomly. Finally, all remaining genes from the genomic, transcriptomic, and proteomic data sources were intersected with the latest list of protein-coding gene symbols (04.08.20) from the HUGO Gene Nomenclature Committee (HUGO) to exclude non-standard gene names.

Gene Occurrence: For every final table of processed data, the top 20 genes were visualized in bar plots, showing the abundance and mean direction of every gene’s regulation. The mean direction of regulation for each gene is represented by the bar plots’ corresponding
The color represents the ratio of studies in which a gene was significantly up- or downregulated. The mean direction of regulation was computed as follows:

Equation 1: Calculation of mean regulation of direction

\[
\text{MeanRegDir(gene)} = \frac{1}{n} \sum_{i=1}^{n} \left( \text{sig(ge}_\text{ne} \text{foldchange}_i) \right)
\]

with:

\[
\text{sig}(x) = \begin{cases} 
1, & \text{if } x > 0 \\
-1, & \text{if } x < 0 
\end{cases}
\]

Intersection: By intersecting the three analyzed omics-layers per disease and the four diseases per omics-layer, it is possible to test if the number of shared genes between some omics-layer or diseases is significantly increased. We used a hypergeometric test to test the overlapping sets, having the total amount of 19,324 Gene Symbols of protein-coding genes (HUGO Gene Nomenclature Committee 04.08.20) (Braschi et al., 2019) as the total population. Intersections were performed and visualized using the R package Venn (Dusa, 2016).

**GO-Term- & Pathway-Analyses:** Independent of the test results, these sets of overlapping genes were also used for KEGG-pathway analyses (M Kanehisa - Novartis Foundation Symposium, 2002) and GO-Term (Gene Ontology Consortium, 2015) analysis. We used the R API WebGestaltR 0.4.4 of the online tool WebGestalt 2020 (Liao et al., 2019) for performing Overrepresentation Analyses (ORA) for all possible intersections per disease and per omics-layer. For performing the ORA, the command WebGestaltR was used with the options:

- `enrichDatabase=c("pathway_KEGG","geneontology_Biological_Process","geneontology_Cellular_Component","geneontology_Molecular_Function")`
- `interestGeneType="gene symbol`
- `referenceSet="genome"
- `topThr = 10000`
- `reportNum = 10000`

The organism was set to "hsapiens" by default.


**Funding:** This research was partially funded by the Emergent AI Center funded by the Carl-Zeiss-Stiftung and the Leibniz Institute for Resilience Research.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.
References


Briston, T. and Hicks, A. R. (2018) ‘Mitochondrial dysfunction and neurodegenerative


Kampinga, Harm H., S. B. (2016) ‘Heat shock proteins as potential targets for protective strategies in neurodegeneration’, *The Lancet Neurology*. Available at:


Kumar, P. et al. (2016) ‘Ion Channels in Neurological Disorders’, in *Advances in Protein


areas of Parkinson’s and Alzheimer’s disease patients’, *Scientific Reports*. Nature Research, 10(1). doi: 10.1038/s41598-020-70174-z.


