Using a Systems Biology Approach To Reveal a Common Molecular Signature Across Neurodegenerative Diseases

By Kareem Fareed

Author Bio

Kareem Fareed is a senior at University School in Hunting Valley Ohio. He is passionate about neuroscience, as well as the application of technology and artificial intelligence in different areas of medicine. He plans to continue his work in these fields and utilize them in a future career as a neurosurgeon.

Abstract

Becoming increasingly prevalent in recent years, the three major neurodegenerative diseases - Alzheimer's (AD) Parkinson's (PD), and Huntington's (HD) - continue to avoid most attempts at treatment and early diagnosis. Operating under the same pathogenic umbrella of neuronal death, it stands to reason that these, and all other neurodegenerative diseases would share risk factors by virtue of their similar pathologies. The aim of this study is to analyze and compare the molecular signatures of the three neurodegenerative diseases, using a holistic approach, in order to uncover a common signature among them. To achieve this, three gene expression datasets are used, one for each disease. A molecular signature of each of the three neurodegenerative diseases is defined using system biology approaches. The resulting signatures are interrogated for any statistically significant overlap between the diseases at different genomic levels. The results of the overlap analysis point to common malfunctions in various autophagic, apoptotic, and mitophagic pathways. Locating a shared signature of the three diseases opens new avenues for drug repositioning across neurodegenerative diseases, offering novel ways of prevention and treatment and supporting future research toward precision medicine.

Keywords: Neurodegenerative, Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Bioinformatics, Systems Biology, Genetics, Convolutional Neural Network, AI, Neuroscience, Precision Medicine, Autophagic pathways, Apoptotic pathways, Mitophagic pathways, Overlap analysis, Genomic levels.

Introduction

Neurodegenerative diseases are a class of diseases marked by the progressive degeneration of neurons, the building blocks of the nervous system. These diseases often manifest as a decline in cognitive and motor function, severely impacting the quality of life for affected individuals. Parkinson's Disease, Alzheimer's Disease, and Huntington's Disease are prominent examples, each with unique pathophysiological features. The irreversible and progressive nature of these diseases makes early diagnosis and intervention crucial, hence requiring advanced informatics like deep learning and systems biology. Despite presenting varying symptoms, these three diseases share a very similar pathogenesis. It stands to reason that these, and all other diseases falling under the umbrella of neurodegeneration would share risk factors by virtue of their similar pathologies.

Parkinson's Disease (PD) is a neurodegenerative disease belonging to a group of diseases called Parkinson's Syndrome. There are two types of PD, hereditary etiology and sporadic etiology. Hereditary etiology is a result of a mutation in a specific gene, whereas sporadic etiology can affect anyone and occurs due to risk factors acquired during a patient's lifetime. PD appears with depigmentation and cell loss in the substantia nigra pars compacta (Figure 1) and the presence of Lewy bodies in the remaining nigral neurons (Bateman 2003). The loss of these dopaminergic neurons in the afflicted brain leads to many of the common symptoms of PD such as akinesia (the absence of voluntary or involuntary movements), rigidity (increased resistance to passive flexion or extension of a joint), and rest tremors (irregular movements in muscles when they are relaxed). With subtly progressing yet dangerous diseases such as PD, early diagnosis is of the utmost importance.

Alzheimer's Disease (AD) is the sixth leading cause of death in the United States and the only one among the top ten that cannot be cured, prevented, or treated. The number of people affected by AD is projected to reach 8.4 million by 2030 and the projected cost associated with dementia is estimated to be $1.1 trillion by 2050 (AA, 2023). Similar to other neurodegenerative diseases, AD progresses silently for many years before the onset of symptoms. The silent preclinical stage of the disease is characterized by the accumulation of amyloid beta (αβ) plaques outside the cell and intracellular neurofibrillary tau tangles in the brain.

Huntington's Disease (HD) is an autosomal dominant neurological condition caused by a single defective gene on chromosome 4 (NCBI). The degeneration of neurons in HD primarily causes motor issues, cognitive decline, and psychiatric problems. Motor symptoms often include jerky and uncontrollable movements known as chorea (Roos 2010). Cognitive decline manifests as difficulty in organizing tasks or coping with new situations, and over time it progresses to severe dementia. Psychiatric problems vary from depression and mood swings to more severe disorders such as obsessive-compulsive disorder and bipolar disorder. Like those mentioned above, HD has no cure. The disease progresses over 10-25 years and ultimately has life-threatening complications.

Neurodegenerative diseases have a complex nature as multiple modifiable, non-modifiable genetic, and non-genetic factors contribute to the diseases. This calls for the use of multiple modalities for neurodegenerative disease detection and prognosis, which are superior to single-modality methods. Autophagy has been shown to be a common factor in the progression of Alzheimer's, Parkinson's, and Huntington's (Guo et al. 2017). Autophagy is the cellular process of degrading or 'recycling' cellular components, and in failing, can lead to one of the above diseases. Such macroscopic-level commonalities tend to make biological sense. However, these biologically-implied similarities aren't easily visible on a more microscopic level.

Many statistical and computational methods have been developed for predicting the development and speed of progression of neurodegenerative diseases from the preclinical stage. Deep learning methods have been at the forefront of such efforts. For example, many Neural Network-based models, like Conventional Neural Networks and generative adversarial networks, have been developed and tested on magnetic resonance images of subjects with early symptoms, like memory concerns or mild cognitive impairment, with varying prediction accuracies (Reas et al. 2023, Qu et al. 2023, Zadeh 2023).

Other methods based on calculating polygenic risk score (PRS) from the subject's genotype data have been used to analyze susceptibility to developing the disease (Wightman et al. 2021). PRS is the product of statistical methods that have proven useful for disease prediction and prognosis, especially when combined with other biomarkers; however, it has limitations. The score calculation is heavily dependent on significant single nucleotide polymorphisms that are defined in the study and do not incorporate rare variants, which limits its ability to capture the whole genomic picture of the disease, especially if the condition is polygenic. Like all complex diseases, synergistic and non-linear interactions must be captured while computing the genetic risk factor of neurodegenerative diseases, which cannot be accounted for using PRS since it assumes linear, additive relationships. Moreover, PRS is more sensitive to the lack of diversity in the data. For example, the majority of available AD data is for subjects of European ethnicity while the disease is more prevalent in Asian, Hispanic, and African ethnicities (Matthews et al. 2018).

Every level of genetic data, going from the finest granularity to the coarsest, gives a different perspective on the underpinnings of the disease. Integrating multi-level omics data (e.g., genomic, transcriptomic, proteomic, epi-genomic) is a promising way to shed light on biological pathways beyond the potential of a single layer of omics data. Thus, the integration of omics data and the use of a holistic systems biology approach is warranted to get a comprehensive view of the underlying mechanisms of neurodegenerative diseases.

Methods

ND Gene Expression Datasets	# Cases	# Controls	# Screened Genes	# DE Genes	Significance Threshold (P-value)
Alzheimer's Disease (ADNI)	270	259	4,952	251	$< 10^{-6}$
Parkinson's Disease (PPMI)	260	256	188	21	$< 10^{-6}$
Huntington Disease (GEO)	109	88	198	15	$< 10^{-4}$

Table 1. Descriptive statistics of the Neurodegenerative Disease (ND) datasets used in this study

Differential expression analysis was performed on each of the three gene expression datasets (Table 1) representing the three diseases under study. Using DESeq2, differentially expressed genes for each disease were chosen for downstream analysis (Love 2014). Differentially expressed genes were filtered at a high statistical significance of false discovery rate (FDR) of 0.2 and minimum fold change (FC) of 2 (p-value < 10^{-2}). Then the genes of each disease were mapped to biological functions using WebGestalt for enrichment analysis of the disease-associated gene sets (Zhang 2005). Biological process terms were used and the FDR threshold was set to 0.2. The resulting gene set for each of the three neurodegenerative diseases was fed into NetworkAnalyst (Xia 2014) for disease-associated tissue-specific Protein-Protein Interaction network analysis and visualization and for pathway analysis using the Kyoto Encyclopedia of Genes and Genomes (Kanehisa 2000). Significant protein interactions (PPI) and functional pathways were filtered at a p-value < 10^{-2} and the FDR threshold was set to 0.2.

NetworkAnalyst is a web-based tool for statistical and network-based meta-analysis of gene expression data that outputs statistically significant PPI networks that harbor the differentially expressed genes from each gene expression dataset. For example, from the AD gene expression dataset, functional enrichment analysis was performed on the differentially expressed genes, and then the resulting gene set was fed into NetworkAnalyst for AD-associated hippocampus-specific PPI network analysis (The brain hippocampus was chosen in this case since it is known to be the first tissue significantly affected by AD). The intersection of the three disease-associated PPI networks, the output of the three neurodegenerative disease datasets, was computed using NetworkAnalyst. The resulting PPI network, at the intersection of the three disease networks, thus represented the common signature across the three neurodegenerative diseases.

Overlap Analysis

To characterize the functional overlap between the biological functions that are identified to be significantly associated with the three neurodegenerative diseases, based on the subjects of the three disease datasets, the pairwise overlap ratio is computed as the fraction of common significant biological function (genes, protein interaction networks, pathways) in every two diseases among the number of significant biological function in the two diseases (i.e., Jaccard coefficient).

Validation

Permutation tests were used to test the statistical significance of the results of the overlap analysis. First, a gene pool was created. For each study, the genotyping array model was identified. All RefSeq transcripts for hg38 assembly were downloaded from the UCSC Table browser and extended 100kb upstream and 10kb downstream and Refseq IDs were translated to HGNC symbols. All HGNC symbols matching Refseq IDs were combined and duplicates were removed. The final HGNC list served as the gene pool for random gene sampling. Gene sets are then randomly sampled such that the size of each set matches that of the observed gene set for each disease. The point of this was to generate a nonsensical dataset that shouldn't have any significant relationships.

A thousand permutated data sets were created with a permuted set size for each population matching the original data set. In each permutation, permuted overlap ratios were calculated as described in the overlap analysis. This resulted in 1000 matrices of pairwise overlap ratios between the three cohorts corresponding to each permuted data set. The mean values of the overlap ratio across all permutations were calculated. These calculations were then used to visualize the overlap in permuted datasets and assess the significance of the observed overlap (p-value < 10-2). The analysis of the permuted datasets should show less overlap as compared to the overlap analysis performed on the original dataset.

Results

The overlap analysis across the three neurodegenerative diseases revealed commonalities at the network level. Figure 1 illustrates the calculated commonalities between the three diseases at two different granularities. The overlap between the three diseases noticeably increased from gene (Figure 1.a) to network analysis (Figure 1.b). For example, there was minimal (1% between AD & HD and 2% between AD & PD) pairwise overlap at the gene level, but the overlap increased at the network level (Figure 1). The network-level analysis captured 23% overlap between AD and PD, 16% overlap between AD and HD, 11% overlap between PD and HD, and 2% overlap between the three, originally disjoint, disease datasets.

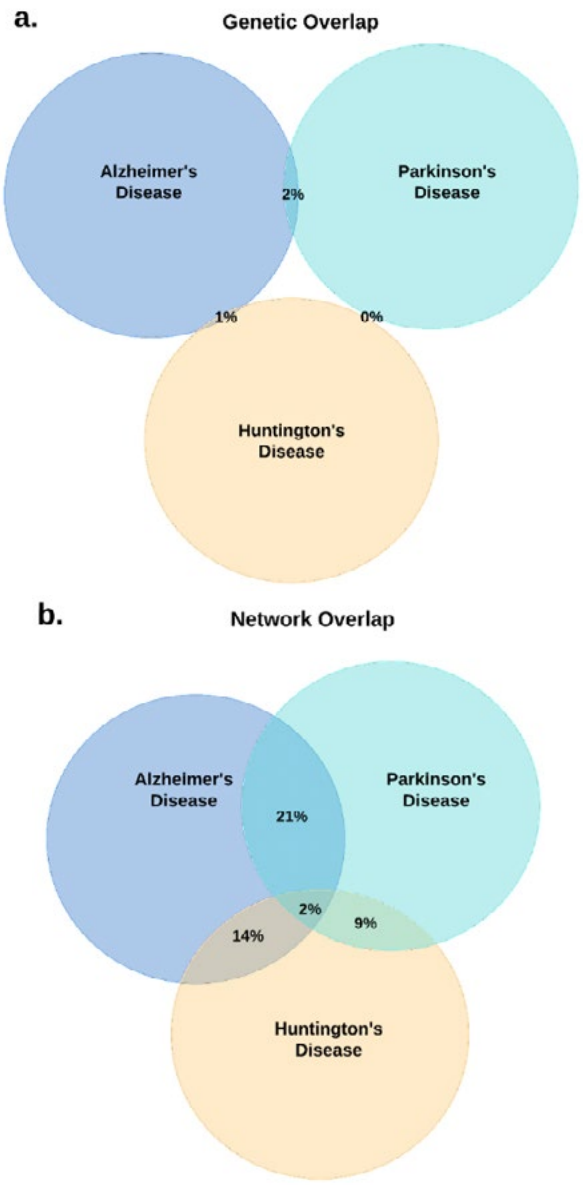

Figure 1. a. represents the Gene-level genetic overlap between the three diseases. b. displays the network-level analysis overlap between the three diseases. This figure highlights that despite displaying very little overlap at the Gene level, network analysis revealed a significant overlap between the three diseases providing a clearer picture of the commonalities between the diseases.

The significance of the overlap was visualized in heatmaps in comparison to their permutated counterparts in Figure 2 (p-value $<10^{-2}$). The permuted dataset (bottom row) showed no significant overlap

among the diseases indicating that the overlap found on the original dataset (top row) was significant. Permutation tests were done on both levels of overlap, genetic (left column) and network (right column).

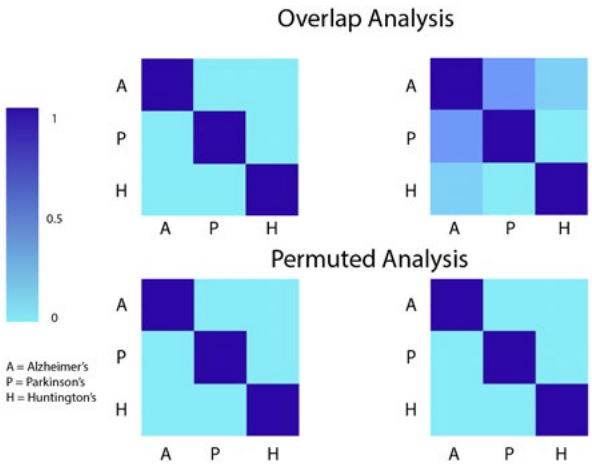

Figure 2. Pairwise overlap between the three neurodegenerative diseases with the gene-level overlap (left) and the network-level overlap (right). The computed pairwise overlaps between the neurodegenerative diseases (n=3). The correlation matrices were n x n matrices of pairwise overlaps between cohorts computed as the Jaccard Coefficient of similarity. The bottom row shows the corresponding results from the permuted data sets, the mean overlaps of a thousand overlaps computed on a thousand permuted datasets. $p<10^{-2}$. The analysis of the permuted datasets shows less overlap as compared to the overlap analysis performed on the original dataset which indicates that the results were significant.

The common neurodegenerative disease PPI is enriched in neurodegenerative disease-associated pathways (Figure 3, Table 2). For example, the MAP3K7 gene (3.a) is enriched in the Autophagy–animal pathway which plays a major role in housekeeping wherein it helps cells remove damaged components and maintain homeostasis. The UBA52 gene (3.b) is enriched in the Mitophagy- animal pathway which eliminates damaged mitochondria. The NFKB1 gene (3.c) is enriched in the Apoptosis pathway which is used to get rid of cells that are either harmful or useless and is crucial to maintaining a functional organism. Finally, the MAPK8 gene (3.d) is enriched in all three pathways.

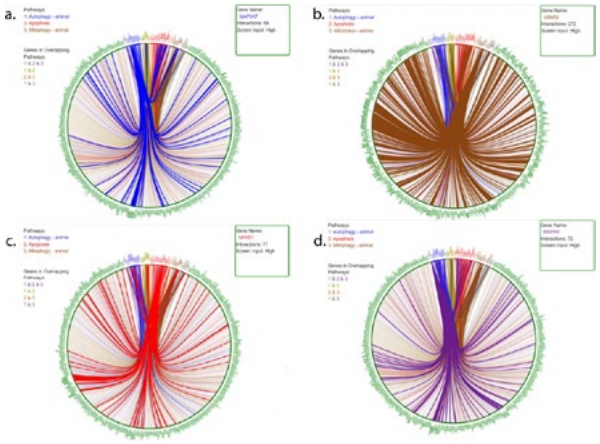

Figure 3. A Circos-visualized example of three significant Neurodegenerative disease pathways: 1. autophagy – animal 2. apoptosis and 3. mitophagy – animal, that were found significant in the datasets of the three diseases (Alzheimer's, Huntington's, Parkinson's) (Krzywinski 2009). The diagram shows interactions between genes enriched in the pathway mitophagy – animal in brown, interactions between genes enriched in the pathway apoptosis in red, and interactions between genes enriched in the autophagy – animal pathway in blue and interacting genes enriched in all selected three pathways in purple. All example genes were significantly enriched in their corresponding pathways and were found significantly expressed in the datasets of the three neurodegenerative diseases.

Pathway	p-value	FDR
Pathways in cancer	5.43×10^{-57}	1.73×10^{-54}
MAPK signaling pathway	9.00×10^{-48}	1.43×10^{-45}
PI3K-Akt signaling pathway	7.85×10^{-43}	8.32×10^{-41}
Focal adhesion	9.92×10^{-41}	7.89×10^{-39}
Ras signaling pathway	3.32×10^{-34}	2.11×10^{-32}
Hippo signaling pathway	1.77×10^{-28}	5.12×10^{-27}
Autophagy - animal	2.45×10^{-19}	3.00×10^{-18}
Apoptosis	5.06×10^{-16}	3.50×10^{-15}
Mitophagy - animal	3.57×10^{-9}	1.06×10^{-8}

Table 2. Pathway Enrichment Analysis on the overlap between the three neurodegenerative diseases (Alzheimer's, Parkinson's, and Huntington's). The table presents a sample of the common pathways that were relevant to the three.

Discussion

This study presents a comprehensive approach to understanding the molecular underpinnings of Alzheimer's, Parkinson's, and Huntington's diseases, using a robust differential expression analysis followed by network-based approaches. The significant findings from our analysis emphasize the complex yet potentially interconnected nature of these neurodegenerative diseases.

A key finding from the study is the minimal overlap at the gene level but a more substantial intersection at the network level between the diseases. This suggests that while the genetic underpinnings might be distinct, these diseases may converge at the network and pathway levels. The permutation tests further validated the significance of these overlaps, suggesting that the observed intersection is not a product of random chance.

Another significant aspect of our findings is the common malfunctions in autophagy, mitophagy, and apoptosis pathways among the three diseases. The central role of the MAP3K7, UBA52, MAPK8, and NFKB1 genes across these pathways suggests its potential as a universal component in the pathogenesis of these conditions. This discovery opens new therapeutic avenues, particularly in the realm of precision medicine.

Conclusion

This study highlights the shared and distinct molecular mechanisms underlying Alzheimer's, Parkinson's, and Huntington's diseases. The results of the overlap analysis revealed the genes MAP3K7, UBA52, MAPK8, and NFKB1 as common markers between the three diseases. They are implicated in autophagy, mitophagy, and apoptosis. Although this paper developed a molecular signature across three neurodegenerative diseases, the analysis could be made more robust through replication analysis using different datasets of the same diseases. The next steps would also be to test this signature against other neurodegenerative diseases like amyotrophic lateral sclerosis (ALS), Lewy-body dementia, and multiple sclerosis (MS) to see if they share the same signature. Then, the molecular signature of the diseases should be used as the target for a drug that can be used against all three diseases. In conclusion, this research opens up possibilities for the development of treatments that could be effective across different neurodegenerative conditions, marking a significant advancement toward more effective and personalized treatment approaches in the future.

References

Bateman, D. (2003). Neurological disorders: course and treatment. 2nd edition. Journal of Neurology, Neurosurgery, and Psychiatry, 74(12), 1700–1700. https://doi.org/10.1136/jnnp.74.12.1700

Alzheimer's Association. (2023). On the Front Lines: Primary Care Physicians and Alzheimer's Care in America. Alzheimer's Association. https://www.alz.org/media/Documents/alzheimers-facts-and-figures.pdf

Center, N. Huntington disease. Nih.gov; National Center for Biotechnology Information (US). https://www.ncbi.nlm.nih.gov/books/NBK22226/

Roos, R. A. C. (2010). Huntington's disease: a clinical review. Orphanet Journal of Rare Diseases, 5(1). https://doi.org/10.1186/1750-1172-5-40

Guo, F., Liu, X., Cai, H., & Le, W. (2017). Autophagy in neurodegenerative diseases: pathogenesis and therapy. Brain Pathology, 28(1), 3–13. https://doi.org/10.1111/bpa.12545

Reas, E. T., Alexey Shadrin, Frei, O., Ehsan Motazedi, McEvoy, L. K., Bahrami, S., Dennis, Makowski, C., Loughnan, R., Wang, X., Broce, I., Banks, S. J., Fominykh, V., Cheng, W., Holland, D., Smeland, O. B., Seibert, T. M., Geir Selbæk, Brewer, J. B., & Fan, C. (2023). Improved multimodal prediction of progression from MCI to Alzheimer's disease combining genetics with quantitative brain MRI and cognitive measures. Alzheimer's & Dementia, 19(11), 5151–5158. https://doi.org/10.1002/alz.13112

Qu, C., Zou, Y., Ma, Y., Chen, Q., Luo, J., Fan, H., Jia, Z., Gong, Q., & Chen, T. (2022). Diagnostic Performance of Generative Adversarial Network-Based Deep Learning Methods for Alzheimer's Disease: A Systematic Review and Meta-Analysis. Frontiers in Aging Neuroscience. https://doi.org/10.3389/fnagi.2022.841696

Zadeh, F. S., Molani, S., Orouskhani, M., Rezaei, M., Shafiei, M., & Abbasi, H. (2023). Generative Adversarial Networks for Brain Images Synthesis: A Review. ArXiv.org. https://arxiv.org/abs/2305.15421

Wightman, D. P., Jansen, I. E., Savage, J. E., Shadrin, A. A., Bahrami, S., Holland, D., Arvid Rongve, Børte, S., Winsvold, B. S., Ole Kristian Drange, Martinsen, A. E., Anne Heidi Skogholt, Willer, C., Geir Bråthen, Ingunn Bosnes, Jonas Bille Nielsen, Fritsche, L. G., Thomas, L. F., Pedersen, L. M., & Gabrielsen, M. E. (2021). A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nature Genetics, 53(9), 1276–1282. https://doi.org/10.1038/s41588-021-00921-z

Matthews, K. A., Xu, W., Gaglioti, A. H., Holt, J. B., Croft, J. B., Mack, D., & McGuire, L. C. (2018). Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged ≥65 years. Alzheimer's & Dementia, 15(1), 17–24. https://doi.org/10.1016/j.jalz.2018.06.3063

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12). https://doi.org/10.1186/s13059-014-0550-8

Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Research. 2005;33(Web Server):W741-W748. doi:https://doi.org/10.1093/nar/gki475

Xia, J., Benner, M. J., & Robert. (2014). NetworkAnalyst - integrative approaches for protein–protein interaction network analysis and visual exploration. Nucleic Acids Research, 42(W1), W167–W174. https://doi.org/10.1093/nar/gku443

M. Kanehisa. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 28(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Krzywinski, M., Schein, J. E., İnanç Birol, Connors, J. M., Gascoyne, R. D., Horsman, D., Steven J.M. Jones, & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. Genome Research, 19(9), 1639–1645. https://doi.org/10.1101/gr.092759.109