S.I. : Genetic pathways to Neurodegeneration

Shared mechanisms among neurodegenerative diseases: from genetic factors to gene networks

Douglas Arneson[1], Yong Zhang[1], Xia Yang[1,*], Manikandan Narayanan[2,3,*]

[1]Department of Integrative Biology and Physiology, University of California Los Angeles, Los Angeles, CA, USA
[2]Systems Genomics and Bioinformatics Unit, Laboratory of Systems Biology, National Institutes of Health, Bethesda, MD, USA
[3]Present address: Department of Computer Science and Engineering (CSE), and Initiative for Biological Systems Engineering (IBSE), Indian Institute of Technology (IIT) Madras, Chennai, TN, India
*Correspondences: nmanik@cse.iitm.ac.in, xyang123@ucla.edu

Unedited version

Abstract

Neurodegenerative diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS) are pressing health concerns in modern societies for which effective therapies are still lacking. Recent high-throughput genomic technologies have enabled genome-scale, multidimensional investigations to facilitate a better understanding of the underlying mechanisms and the identification of novel targets. Here we review the molecular insights gained through such studies, and compare the similarities and differences between neurodegenerative diseases revealed by systems genomics and gene network modeling approaches. We focus specifically on the shared mechanisms at multiple molecular scales ranging from genetic factors to gene expression to network-level features of gene regulation, and whenever possible also point out mechanisms that distinguish one disease from another. Our review sets the stage for similar genome-wide inspection in the future on shared/distinct features of neurodegenerative diseases at the levels of cellular, proteomic or epigenomic signatures, and how these features may interact to determine the progression and treatment response of different diseases afflicting the same individual.

Introduction

Common neurodegenerative diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), and Huntington's disease (HD) are debilitating disorders with increasing prevalence in modern aging societies. Decades of research on individual diseases via diverse approaches have offered deep insights into the phenotypic manifestations and molecular underpinnings of each disease. Interestingly, patterns of converging features across neurodegenerative diseases such as dementia in AD, PD, and ALS have been observed(PARIKSHAK *et al.* 2015; LYNCH *et al.* 2016; SANTIAGO *et al.* 2017), calling for a quest to better understand the relationships between neurodegenerative diseases to reveal disease-specific mechanisms as well as potentially shared mechanisms. While disease-specific mechanisms are useful for personalized, disease-specific therapies, the shared mechanisms revealed can facilitate the development of common therapies for multiple neurodegenerative diseases and a better understanding of the co-occurrence and overall interactions among these diseases.

While disease-disease overlap may be expected due to shared environmental risk factors such as common chemical exposures between different diseases(LIU *et al.* 2009), recent investigations of the shared mechanisms or interactions between diseases at the levels of individual risk factors (genetic vs. environment), candidate disease genes, pathways, and networks have all supported the presence of shared molecular pathways(PARIKSHAK *et al.* 2015; LYNCH *et al.* 2016; SANTIAGO *et al.* 2017). These studies have provided valuable information on common dysregulation in different diseases, but were done at various times in the past on different subsets of diseases, and a systematic review of the most current information is needed to achieve a more comprehensive understanding of the shared mechanisms across neurodegenerative diseases.

In this review, we systematically collect multidimensional molecular information probing the mechanisms of individual diseases at genetic, transcriptome, pathway, and gene network levels. We focus on genome-wide studies to offer an objective and comprehensive view, and summarize the shared mechanisms between diseases revealed at each molecular level to derive more conclusive insights. Here we focus on AD, PD, ALS, and HD whenever sufficient data is available. We organize our review of the genome-wide similarities among neurodegenerative diseases based on the scale of investigation, that is, whether the shared (and at times unique) factors discussed are at the genetic, gene expression or network level.

Shared genetic factors between neurodegenerative diseases

Neurodegenerative disorders all have significant genetic components, with genetic heritability for AD, PD, and ALS estimated to be 60-80%(GATZ *et al.* 2006; VAN CAUWENBERGHE *et al.* 2016), ~40%(HAMZA AND PAYAMI 2010), and ~60%(AL-CHALABI *et al.* 2010), respectively. Mendelian forms of neurodegenerative diseases have been attributed to rare mutations, such as *APP* (amyloid precursor protein) and presenilin genes (*PSEN1*, *PSEN2*) for AD, *SNCA* (α-synuclein), *PARK2* (Parkin), *PINK1* (PTEN-induced putative kinase I), *MAPT* (microtubule-associated protein tau) and *LRRK2* (leucine-rich repeat kinase 2 or dardarin) for PD, *SOD1* (cytosolic Cu/Zn superoxide dismutase), *ALS2* (alsin), *SETX* (senataxin) and *VAPB* (synaptobrevin/ VAMP (vesicle-associated membrane protein)-associated protein B) for ALS(BERTRAM AND TANZI 2005; PASINELLI AND BROWN 2006), and *C9ORF72* (chromosome 9 open reading frame 72) repeat expansion as a shared genetic cause for ALS and FTD(VAN BLITTERSWIJK *et al.* 2012) which was dissected recently using CRISPR-Cas9 screens(KRAMER *et al.* 2018). However, the majority of disease cases are of non-Mendelian form and exhibit complex etiology involving large numbers of genetic variants with moderate to subtle effects. These complex forms of neurodegenerative diseases are difficult to examine, until the advent of genome-wide association studies (GWAS).

Within a decade of the best known early GWA study in 2005(KLEIN *et al.* 2005), the GWAS approach has become a staple in modern genetic research and a driving force for novel causal gene discoveries for numerous human traits and diseases including neurodegenerative diseases. Based on the recent GWAS catalog(WELTER *et al.* 2014) (https://www.ebi.ac.uk/gwas/; downloaded on Sep 19, 2017), a total of 43, 26, and 17 studies have been conducted for AD, PD, and ALS, revealing tens of single nucleotide polymorphisms (SNPs) to be associated with each of these diseases at genome-wide significance ($p<5x10^{-8}$; HD was not included in the analysis due to the few limited loci identified) (**Figure 1**). Among these, the genetic locus with the strongest effect and clinical significance for AD is the *APOE* locus, which has also been implicated in PD (albeit with different alleles). Direct comparison of the significant GWAS hits between diseases revealed no overlap in SNPs (**Figure 1A**), but a few overlaps in the candidate genes mapped to the top SNPs (**Figure 1B**) as well as in the over-represented pathways among the

Uncorrected Version

candidate genes (**Figure 1C**). Specifically, *HLA-DRB5* and *MAPT* were GWAS candidate genes for both AD and PD. At the pathway level, "vesicle-mediated transport" was shared across all 3 diseases, and 9 pathways such as synaptic signaling, neuron projection development, and proteolysis were shared between AD and PD (**Table 1**).

Previously, Ramanan and Saykin(RAMANAN AND SAYKIN 2013) have categorized 13 and 15 GWAS candidate genes for AD and PD, respectively, based on the known functions of individual genes and compared the convergent pathways between the two diseases. The genes included represented a much smaller number of candidate genes than what was summarized above in our analysis. In the earlier analysis by Ramanan and Saykin, as long as a pathway was implicated by a single gene, the pathway was considered to be implicated in the disease biology. Using this strategy, they identified numerous shared pathways between AD and PD. These included intracellular processes (apoptosis, autophagy, mitochondrial function, oxidative damage/repair, proteasome), pathways involving local tissue environment (cell adhesion, endocytosis, neurotransmission, prions/transmissible factors), pathways related to systemic environment (inflammation/immune system, lipid/metabolic/endocrine, vascular factors), and processes relevant to development and aging (epigenetics, neurotrophic factors, telomeres). Our current analysis based on the over-represented pathway among the updated GWAS candidates confirmed the involvement of both intracellular and intercellular processes (multicellular organismal process, transport, macromolecular complex binding, cell activation, vesicle mediated transport, proteolysis, kinase binding), and also implicated pathways associated with the nerve system (synaptic signaling, neuron projection) (**Table 1**). Interestingly, Ramanan and Saykin also attempted to model the interactions among the AD/PD candidate genes using transcription factor binding networks curated in the MetaCore software. They found that 9 of the 13 AD GWAS genes and 10 of the 15 PD GWAS genes were tightly connected within a coherent network coordinated by transcription factors SP1 and AP-1.

The above analyses support the presence of shared GWAS genes, pathways, and networks between several neurodegenerative diseases. However, these were based on the significant GWAS loci only. Given that ample evidence supports that complex traits or diseases involve numerous genetic variants with effect sizes ranging from strong to moderate to subtle, the top-loci focused comparisons miss the opportunity to obtain comprehensive mechanistic insights. Additionally, examining overlaps in the top loci between diseases would necessitate the assumption that the ranks of genetic association strengths for the GWAS loci are similar across diseases, which may not be true in that strong loci for one disease may only subtly perturb another disease. Therefore, it is important to compare the disease mechanisms revealed by GWAS using the full GWAS statistics. As an illustration of this concept, we took the significant GWAS SNPs for PD and ALS from GWAS Catalog and plotted their association p values with AD based on the full summary statistics from the IGAP GWAS study(LAMBERT *et al.* 2013) on AD using Q-Q plots (**Figure 1D-E**) to assess whether the observed distribution of p values is significantly different from the expected null distribution. A significant deviation of the AD association p values among the PD or ALS SNPs from the null distribution towards lower p value (higher -log10(p)) values would indicate that PD or ALS SNPs collectively show stronger association with AD than random SNPs, which can serve as evidence for genetic sharing. This analysis showed that top PD SNPs are collectively associated with AD but in a subtler way, as the AD GWAS p values for PD SNPs were mostly less significant than 10^{-3}. In contrast, the top ALS SNPs did not appear to show evidence of association with AD.

Recently, multiple more sophisticated statistical and bioinformatics methods have been used to explore shared genetic components across diseases(FORTUNE *et al.* 2015; BROWN *et al.*

Unedited version

2016; PICKRELL *et al.* 2016; SHU *et al.* 2017). The various methods can detect genetic sharing at SNP, gene, pathway, and network levels. For example, Pickrell et al.(PICKRELL *et al.* 2016) examined the overlapping genetic architecture among 43 human traits, including AD and PD. They used a log likelihood-based model selection method and revealed genetic sharing between many diseases. In particular, their study also provided supporting evidence for moderate genetic sharing between AD and PD, agreeing with our analysis above (ALS was not examined in their analysis). Future applications of these various methods will further our understanding of genetic sharing between neurodegenerative diseases.

Shared gene expression signatures

With the advent of microarrays and next generation RNA sequencing, there has been an outpour of studies on transcriptional profiling of many complex diseases to better understand the underlying disease mechanisms. Genome-wide gene expression profiles collected from disease vs. control individuals represent the largest publicly available resource in the genomic domain. As such, there have been numerous studies which have profiled the transcriptomes of different neurodegenerative diseases and explored the corresponding overlapping gene signatures to determine the shared transcriptional perturbations of neurodegenerative diseases. In addition to these numerous individual profiling studies, there have also been efforts to boost the power of detection and find consistent gene signatures by conducting meta analyses and comparing two or more neurodegenerative diseases to find conserved gene perturbations.

Within the field of neurodegenerative diseases, it has been well established that there is a higher probability of developing concurrent PD and AD than would be expected by random chance, making this disease comparison a prime candidate for the exploration of overlapping gene signatures. This was precisely the goal in a comparative study profiling three brain regions (hippocampus, gyrus-frontalis-medius, and cerebellum)(GRÜNBLATT *et al.* 2007). This study found 12 genes that were similarly perturbed by PD and AD across all three brain regions including: synaptic vesicle genes (*SYT1*), Alzheimer's related-genes (*APP, SNX2*), insulin genes (*IRS4*), and oxidative stress genes (*GSTM1*). In addition, the study identified 4 genes (*CNR2*, *HIST1H3E*, *CHRNA6*, and *BACE1*) which showed opposite regulation patterns between the two diseases. For instance, BACE1, which is involved in processing Amyloid precursor protein, was found to be upregulated in PD but downregulated in AD(GRÜNBLATT *et al.* 2007).

While the previous study was based on data-driven whole transcriptome profiling to find similarities between PD and AD, there have been more targeted efforts to find consistencies in transcriptomes between neurodegenerative diseases. Specifically, one study drew on prior literature evidence of inflammation and perturbation of the immune system, which is characteristic of AD, PD and Creutzfeldt-Jakob disease, to find transcriptomic overlaps. Using this approach, Gonzalez *et. al* found overlaps between cytokines and the mediators of the immune response in all three diseases, as well as additional overlaps between AD and PD characterized by inflammatory markers in the blood and serum(LÓPEZ GONZÁLEZ *et al.* 2016).

Although there are many more examples of targeted gene expression profiling studies which serve to characterize a single neurodegenerative disease or compare between two diseases, the most comprehensive gene signature comparison between neurodegenerative diseases to date is a meta-analysis of 1,270 post-mortem brain tissues from 13 patient cohorts spanning four neurodegenerative diseases (AD, PD, HD, and ALS) across many different brain regions(LI *et al.* 2014). Sampled tissues include: hippocampus, frontal cortex, entorhinal cortex, dorsolateral prefrontal cortex and medial temporal lobe for AD; substantia nigra, dorsolateral prefrontal cortex, putamen, dorsal motor nucleus and globus pallidus interna for PD; motor cortex,

Unedited Version

ventral head of the caudate nucleus and dorsolateral prefrontal cortex for HD; and motor cortex and cervical spinal cord for ALS (**Table 1**). This meta-analysis identified a shared gene expression signature of 243 genes, which was validated in an additional withheld dataset comprising 205 samples from 15 different cohorts (**Figure 2**). This shared gene signature contains genes associated with bioenergetic deficits, M1-type microglial activation and gliosis, thereby supporting these processes as consistent themes across neurodegeneration(LI *et al.* 2014). Furthermore, pathway enrichment of the differentially expressed genes largely overlapped with literature on known neurodegenerative disease pathways, and included functional processes such as inflammation, altered synaptic transmission, mitochondrial dysfunction, and oxidative stress.

The overlapping gene signatures of ALS, PD, and AD were further explored in another review of neurodegenerative diseases. The authors found overlaps between all three diseases in the form of neuroinflammation gene signatures (consistent with the targeted findings above), with further overlaps identified in genes related to RNA splicing and protein turnover between ALS and PD, and mitochondrial dysfunction genes as a common theme between PD and AD(COOPER-KNOCK *et al.* 2012).

Given that genome-wide gene expression profiling is the most abundant omics-based profiling platform, we focused here on transcriptome-based signature comparisons between neurodegenerative diseases. However, there have been a handful of studies on other data modalities such as shared protein dysregulation(HOSP *et al.* 2015) and shared epigenomic patterns(URDINGUIO *et al.* 2009; PORTELA AND ESTELLER 2010; SANCHEZ-MUT *et al.* 2016). These additional data modalities can provide unique information that is not captured at the transcriptome level and, when more comprehensively investigated in the future, can be integrated into system-wide network models to better capture the shared etiology between neurodegenerative diseases.

Shared network-level dysregulation

Genes and proteins do not work in isolation within the cells and tissues of our body, but instead interact with and regulate each other via protein-protein, protein-DNA, and other biomolecular interactions. Disease-induced perturbations can propagate through this interconnected network of genes and proteins. Evidence for this type of network-level dysregulation shared between different neurodegenerative diseases comes from several studies on transcriptional networks or protein interaction networks. These studies typically infer a transcriptional network by adding a network link between any pair of genes that show correlated expression (coexpression) in postmortem brain samples of a group of individuals (such as a group of individuals with disease, or another group of control individuals without dementia); and typically infer a protein interaction network from more direct experimental assays done on human cells (such as a yeast two-hybrid assay to detect protein-protein interactions, or a ChIP-seq experiment to detect protein-DNA interactions).

Different neurodegenerative diseases show similarity at the level of transcriptional networks. We have found for instance(NARAYANAN *et al.* 2014) that the global transcriptional network in the human dorsolateral prefrontal cortex is drastically altered in both AD and HD diseases in a similar fashion when compared to control individuals. This result is based on a systematic differential coexpression (DC) analysis that revealed two types of disrupted gene-gene relations based on whether the correlation strength between two genes are increased (Gain of Correlation or GOC) or decreased (Loss of Correlation or LOC) in the disease group relative to controls. The network of shared DC relations between AD and HD contained a majority of LOC

Unedited version

relations, even when the individual DC networks (i.e., AD vs. Controls DC network, or HD vs. Controls DC network) contained predominantly GOC relations (**Figure 3A**).

Aligning the shared DC network between AD and HD (which comprised 8043 DC relations involving 3021 genes) with a network of 116,220 protein-protein/protein-DNA interactions from HPRD, BIND and other databases resulted in a 242-gene subnetwork. This 242-gene subnetwork was enriched for independent AD and HD signatures and for several biological processes (including neuron differentiation, gap junction trafficking, regulation of apoptosis and other processes; **Table 1**). This subnetwork also revealed two interacting processes involving GOC of chromatin organization genes and LOC of oligodendrocyte differentiation genes as a pathological mechanism shared between AD and HD(NARAYANAN *et al.* 2014).

We were interested in comparing genes in the shared HD-AD DC network to genes in the coexpression network of other neurodegenerative diseases. The most suitable such study in terms of sufficient sample size to construct robust coexpression networks was a study on ALS(SARIS *et al.* 2009). However, due to the lack of access to brain tissues, the ALS study was based on blood gene expression analysis. Using blood expression data from multiple ALS cohorts, this study identified and replicated significant associations of two sets of coherently regulated genes (or coexpression modules) with ALS status: a blue module enriched for genes upregulated in ALS patients compared to controls, and a yellow module enriched for genes downregulated in ALS. In a test for enrichment of the top 500 genes correlated to the activity of the ALS yellow (or blue) modules for various functional or disease gene categories, the study found significant overlap with HD disease genes category besides other neurological disorder categories such as atrophy of dendrites. Furthermore, the top 500 blue or yellow module genes (i.e., most positively correlated 500 genes averaged across multiple ALS cohorts as reported in in their Supplementary Table S4) overlapped significantly with the genes in the shared DC network between HD and AD in our study (**Figure 3C**). These results taken together support significant sharing of coexpression network-level signatures between HD, AD and ALS.

We focused above on transcriptional coexpression network analysis of bulk tissues such as postmortem brain samples or blood samples since they were available from multiple neurodegenerative diseases. There have also been cell type specific studies such as coexpression network analysis of spinal motor neurons implicated in ALS(HO *et al.* 2016) or fibroblasts affected in ALS(KOTNI *et al.* 2016). More such work is needed to understand if different neurodegenerative diseases could affect specific cell types in a shared fashion.

To complement studies that use transcriptional coexpression networks inferred from expression data to uncover disease dysregulation, one could also use literature-based network data on protein-protein, protein-DNA and other physical interactions to dissect disease dysregulation. For instance, a study used a human protein-protein interaction network to expand an initial seed set of 10 common disease susceptibility genes shared between AD, HD and PD (e.g., *ESR2, PARP1, GSK3B, UCHL1* and *LRRK2*, as identified from genetic association databases, literature mining or other sources) into a larger set of 1294 genes connected to these seed genes in the protein network(LI *et al.* 2015). Inspection of the protein network among these expanded genes revealed enrichment of metabolic pathways from the KEGG database, and modules/pathways that provide bridging interactions between the common susceptible genes for these three neurodegenerative diseases. The discovered bridging pathways such as *adherens* and *tight junctions* were further validated using independent gene expression data related to these neurodegenerative diseases.

Unedited version

Given the complexity of how these diseases progress, instead of using gene expression data in the validation phase as done in the study above, one may use it directly in the discovery phase to identify and prioritize the bridging pathways or common disease subnetworks. This approach was taken in a related study(LIU *et al.* 2012) to search for similar subnetworks dysregulated in both AD and PD. Instead of simply looking at first-level neighbors of seed disease genes in the human protein network, this study also employed an objective function based on Steiner trees to expand the seed genes minimally into a connected disease subnetwork, and then inspected the resulting subnetworks for coordinated transcriptional dysregulation in both AD and PD. The reconstructed AD and PD networks respectively contained 225 genes (387 interactions) and 273 genes (502 interactions), with 72 genes shared between the two networks. While the majority of these 72 genes could be constructed from a direct overlap between the two starting seed disease gene sets (derived based on AD or PD association in at least four publications), the power of a network-based approach is in its depiction of the interconnected nature of these shared genes and the extra bridging genes that the Steiner tree algorithm adds. For instance, the Steiner tree algorithm added five AD genes (*ACHE, APP, ATXN1, CLU* and *DAPK1)* to the PD network, and five PD genes (*APOB, CALR, CAV1, NOS1* and *TFRC*) to the AD network based on molecular interactions between the AD and PD genes.

Conclusion

We set out to obtain a systematic genome-wide elucidation of common molecular pathways to neurodegeneration, and did so using three different genomic data types for which data is relatively abundant to allow between-disease comparisons: genetic, gene expression and network-level signatures. The three data types seem to complement one another as the common dysregulated pathways supported by each data type are generally distinct from those supported by the other data types (**Table 1**), with some exceptions like the vesicle mediated transport receiving support from both genetic and gene expression data for its association with AD, PD and ALS. Certain previously known shared mechanisms like protein misfolding and subsequent neuronal loss are only captured partially in our results (i.e., cell death but not protein misfolding is found enriched; **Table 1**). Future work could review other lines of evidence such as shared protein, cellular (composition of cell types) or structural/neuronal network signatures among the neurodegenerative diseases when data becomes more enriched in these domains, and these could for instance reveal both protein misfolding pathways and other novel pathways shared among neurodegenerative diseases.

Several factors need to be considered when interpreting the results in this study. First of all, the biological implications of the different types of data can be different. For instance, genetic information deals with heritable DNA alterations that precede disease development, which may carry stronger implications for disease causality. Gene expression data, on the other hand, are more subject to dynamic changes and can reflect on both causal genes/pathways that are upstream of disease development and reactive genes/pathways that are downstream of disease processes. Therefore, one cannot directly take the large numbers of genes and pathways revealed from gene expression studies as evidence for disease causality. Sophisticated integrative analysis of the various data modalities and perturbation experiments are needed to differentiate causal driver processes from reactive passenger pathways. Second, different statistical significance or other filtering cutoffs were employed in different studies and data modalities to build their disease gene sets – hence it is easier to interpret similarities than differences observed between diseases' gene sets, since differences could simply arise from usage of different cutoffs or other technical/methodological artifact than real differences between the diseases being compared. We note here that we did keep the factors under our control to be

Unedited version

the same across diseases to minimize artefactual results (in **Table 1** for instance, the same pathway definitions and the same pathway enrichment tool based on hypergeometric test was used). A further consideration when comparing gene signatures (and other omics modalities) is the tissue-specific nature of disease signatures as seen in neurodegenerative diseases like HD(HODGES *et al.* 2006) and tissue-specific patterns of splicing(TWINE *et al.* 2011). Therefore, it is important to match the tissue types in which the genes and pathways manifest when comparing mechanisms between diseases. If these tissue-specific influences are not accounted for when comparing diseases, spurious conclusions will result. Having said that, the tissues of relevance to neurodegenerative diseases are predominantly different regions of the brain, with the cortical region profiled in many studies for instance, and therefore allows us to compare different tissue-based disease signatures. Lastly, although we tried to be as comprehensive as possible in collecting data and studies, it is easily discernible that not all comparisons are possible due to missing information or data on one or more diseases. Therefore, future coordinated efforts are needed to systematically collect multi-omics data types from multiple tissues and cell types from multiple neurodegenerative diseases to enable more comprehensive investigations.

In summary, we reviewed common mechanisms shared between neurodegenerative diseases here (i.e., overlap between dysregulation signatures identified by typically studying each disease separately in independent studies or cohorts), and found several pathways to be enriched for genes associated with multiple neurodegenerative diseases. This short review of *disease-disease overlap* could pave the way to understand *disease-disease interactions*, i.e., interaction between two neurodegenerative diseases afflicting the same individual. This is an underexplored topic of research that could constitute a promising avenue for future research, and thereby help reveal new therapies for diseases with mechanistic overlaps and interactions.

Acknowledgment

This work was partially supported by the Intramural Program of NIAID (National Institute of Allergy and Infectious Diseases) at the National Institutes of Health (NIH) to M.N., NIH grants DK104363 and NS103088 to X.Y., and UCLA Hyde Fellowship to D.A.

References

Al-Chalabi, A., F. Fang, M. F. Hanby, P. N. Leigh, C. E. Shaw *et al.*, 2010 An estimate of amyotrophic lateral sclerosis heritability using twin data. J Neurol Neurosurg Psychiatry 81**:** 1324-1326.

Bertram, L., and R. E. Tanzi, 2005 The genetic epidemiology of neurodegenerative disease. J Clin Invest 115**:** 1449-1457.

Brown, B. C., C. J. Ye, A. L. Price, N. Zaitlen and A. G. E. N. T. D. Consortium, 2016 Transethnic Genetic-Correlation Estimates from Summary Statistics. Am J Hum Genet 99**:** 76-88.

Cooper-Knock, J., J. Kirby, L. Ferraiuolo, P. R. Heath, M. Rattray *et al.*, 2012 Gene expression profiling in human neurodegenerative disease. Nat Rev Neurol 8**:** 518-530.

Fortune, M. D., H. Guo, O. Burren, E. Schofield, N. M. Walker *et al.*, 2015 Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. Nat Genet 47**:** 839-846.

Gatz, M., C. A. Reynolds, L. Fratiglioni, B. Johansson, J. A. Mortimer *et al.*, 2006 Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 63**:** 168-174.

Unedited version

Grünblatt, E., N. Zander, J. Bartl, L. Jie, C. M. Monoranu *et al.*, 2007 Comparison analysis of gene expression patterns between sporadic Alzheimer's and Parkinson's disease. J Alzheimers Dis 12**:** 291-311.

Hamza, T. H., and H. Payami, 2010 The heritability of risk and age at onset of Parkinson's disease after accounting for known genetic risk factors. J Hum Genet 55**:** 241-243.

Ho, R., S. Sances, G. Gowing, M. W. Amoroso, J. G. O'Rourke *et al.*, 2016 ALS disrupts spinal motor neuron maturation and aging pathways within gene co-expression networks. Nat Neurosci 19**:** 1256-1267.

Hodges, A., A. D. Strand, A. K. Aragaki, A. Kuhn, T. Sengstag *et al.*, 2006 Regional and cellular gene expression changes in human Huntington's disease brain. Hum Mol Genet 15**:** 965-977.

Hosp, F., H. Vossfeldt, M. Heinig, D. Vasiljevic, A. Arumughan *et al.*, 2015 Quantitative interaction proteomics of neurodegenerative disease proteins. Cell Rep 11**:** 1134-1146.

Klein, R. J., C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler *et al.*, 2005 Complement factor H polymorphism in age-related macular degeneration. Science 308**:** 385-389.

Kotni, M. K., M. Zhao and D. Q. Wei, 2016 Gene expression profiles and protein-protein interaction networks in amyotrophic lateral sclerosis patients with C9orf72 mutation. Orphanet J Rare Dis 11**:** 148.

Kramer, N. J., M. S. Haney, D. W. Morgens, A. Jovicic, J. Couthouis *et al.*, 2018 CRISPR-Cas9 screens in human cells and primary neurons identify modifiers of C9ORF72 dipeptide-repeat-protein toxicity. Nat Genet.

Lambert, J. C., C. A. Ibrahim-Verbaas, D. Harold, A. C. Naj, R. Sims *et al.*, 2013 Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 45**:** 1452-1458.

Li, M. D., T. C. Burns, A. A. Morgan and P. Khatri, 2014 Integrated multi-cohort transcriptional meta-analysis of neurodegenerative diseases. Acta Neuropathol Commun 2**:** 93.

Li, P., Y. Nie and J. Yu, 2015 An Effective Method to Identify Shared Pathways and Common Factors among Neurodegenerative Diseases. PLoS One 10**:** e0143045.

Liu, Y., M. Koyutürk, S. Maxwell, Z. Zhao and M. R. Chance, 2012 Integrative analysis of common neurodegenerative diseases using gene association, interaction networks and mRNA expression data. AMIA Jt Summits Transl Sci Proc 2012**:** 62-71.

Liu, Y. I., P. H. Wise and A. J. Butte, 2009 The "etiome": identification and clustering of human disease etiological factors. BMC Bioinformatics 10 Suppl 2**:** S14.

López González, I., P. Garcia-Esparcia, F. Llorens and I. Ferrer, 2016 Genetic and Transcriptomic Profiles of Inflammation in Neurodegenerative Diseases: Alzheimer, Parkinson, Creutzfeldt-Jakob and Tauopathies. Int J Mol Sci 17**:** 206.

Lynch, M. A., O. Hardiman, M. Elamin, J. Kirby and L. P. Rowland, 2016 Common Themes in the Pathogenesis of Neurodegeneration, pp. 1-12 in *Neurodegenerative Disorders: A Clinical Guide*, edited by O. Hardiman, C. P. Doherty, M. Elamin and P. Bede. Springer International Publishing, Cham.

Narayanan, M., J. L. Huynh, K. Wang, X. Yang, S. Yoo *et al.*, 2014 Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases. Mol Syst Biol 10**:** 743.

Parikshak, N. N., M. J. Gandal and D. H. Geschwind, 2015 Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat Rev Genet 16**:** 441-458.

Unedited Version

Pasinelli, P., and R. H. Brown, 2006 Molecular biology of amyotrophic lateral sclerosis: insights from genetics. Nat Rev Neurosci 7**:** 710-723.

Pickrell, J. K., T. Berisa, J. Z. Liu, L. Ségurel, J. Y. Tung *et al.*, 2016 Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet 48**:** 709-717.

Portela, A., and M. Esteller, 2010 Epigenetic modifications and human disease. Nat Biotechnol 28**:** 1057-1068.

Ramanan, V. K., and A. J. Saykin, 2013 Pathways to neurodegeneration: mechanistic insights from GWAS in Alzheimer's disease, Parkinson's disease, and related disorders. Am J Neurodegener Dis 2**:** 145-175.

Sanchez-Mut, J. V., H. Heyn, E. Vidal, S. Moran, S. Sayols *et al.*, 2016 Human DNA methylomes of neurodegenerative diseases show common epigenomic patterns. Transl Psychiatry 6**:** e718.

Santiago, J. A., V. Bottero and J. A. Potashkin, 2017 Dissecting the Molecular Mechanisms of Neurodegenerative Diseases through Network Biology. Front Aging Neurosci 9**:** 166.

Saris, C. G., S. Horvath, P. W. van Vught, M. A. van Es, H. M. Blauw *et al.*, 2009 Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. BMC Genomics 10**:** 405.

Shu, L., K. H. K. Chan, G. Zhang, T. Huan, Z. Kurt *et al.*, 2017 Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. PLoS Genet 13**:** e1007040.

Twine, N. A., K. Janitz, M. R. Wilkins and M. Janitz, 2011 Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. PLoS One 6**:** e16266.

Urdinguio, R. G., J. V. Sanchez-Mut and M. Esteller, 2009 Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. Lancet Neurol 8**:** 1056-1072.

van Blitterswijk, M., M. DeJesus-Hernandez and R. Rademakers, 2012 How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: can we learn from other noncoding repeat expansion disorders? Curr Opin Neurol 25**:** 689-700.

Van Cauwenberghe, C., C. Van Broeckhoven and K. Sleegers, 2016 The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet Med 18**:** 421-430.

Welter, D., J. MacArthur, J. Morales, T. Burdett, P. Hall *et al.*, 2014 The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42**:** D1001-1006.

Unedited version

Received 13 February; revised 19 March 2018; accepted 20 March 2018

Table 1. Shared molecular pathways between neurodegenerative diseases based on genetic, gene expression, and network evidence. The enrichments marked with an asterisk (*) are taken from prior literature, and other enrichments are representative results from hypergeometric-distribution-based enrichment tests performed by us for this study using a background of 16,954 genes that represent the union of all genes in all pathways considered. All enrichments called at 5% false discovery rate (FDR) and corresponding input disease gene sets are available as two Supplementary Datasets. The N/A in the table means no gene sets were readily available to perform enrichment tests for the corresponding disease and data type.

Diseases	Genetics	Gene Expression	Network
AD/PD/ALS/HD	N/A	• (Tissues^{+}: ALS – motor cortex and cervical spinal cord; HD – motor cortex, ventral head of the caudate nucleus, and dorsolateral prefrontal cortex; PD – substantia nigra, dorsolateral prefrontal cortex, putamen, dorsal motor nucleus, and globus pallidus interna; AD – hippocampus, frontal cortex, entorhinal cortex, dorsolateral prefrontal cortex, and medial temporal lobe) • n = 243 genes(Li *et al.* 2014) • Bioenergetic deficits* • M1-type microglial activation* • Gliosis* • Neuroinflammation* • Immune response/ inflammation • n = 322 genes (Figure 2) • Synaptic signaling • Metabolic deficits • Oxidative Phosphorylation	N/A
AD/ALS/HD	N/A	• (same study and tissues as first entry marked Tissues^{+}) • n = 786 genes (Figure 2) • Cell-cell communication • Immune response/ inflammation	• (Tissues: HD and AD – dorsolateral prefrontal cortex; ALS – blood) • n = 149 genes (union of the 76 and 73 overlapping genes in Figure 3B) • RNA binding • Protein localization • Metabolic processes • Chromatin modification • Regulation of cell death

Unedited version

AD/ALS/PD	• n = 91 AD genes, 76 PD genes, and 33 ALS genes (Figure 1B) • **Vesicle mediated transport**	• (same study and tissues as first entry marked Tissues^{+}) • n = 780 genes (Figure 2) • **Vesicle mediated transport** • Immune response/ inflammation • Regulation of cell death • Cell-cell communication	N/A
AD/PD/HD	N/A	• (same study and tissues as first entry marked Tissues^{+}) • n = 1524 genes (Figure 2) • Protein localization • Immune response/ inflammation • Phosphorylation • Cell death	N/A
ALS/HD/PD	N/A	• (same study and tissues as first entry marked Tissues^{+}) • n = 1019 genes (Figure 2) • Protein localization • Cell death • Neurogenesis • Phosphorylation • Metabolic processes	N/A
AD/PD	• n = 91 AD genes and 76 PD genes (Figure 1B; only 2 gene-level overlaps between these two genesets, but many pathway-level overlaps as shown next) • Negative regulation of multicellular organismal process • Regulation of transport • Synaptic signaling • Macromolecular complex binding • Positive regulation of cell activation	• (Tissues: AD – hippocampus, prefrontal cortex, Brodmann areas, parietal lobe, temporal cortex, cerebellum, frontal lobe, cingulate, occipital cortex, basal ganglia, blood; PD – substantia nigra, frontal gyrus, putamen, cerebellum, occipital cortex, blood) • n = 218 PD and 1178 AD genes (Note: these two genesets are from literature and reported pathways(COOPER-KNOCK *et al.* 2012) consistent between the two genesets are shown next) • Cytokines* • Immune response*	N/A

Unedited Version

	• Regulation of neuron projection development • Regulation of vesicle mediated transport • Positive regulation of proteolysis • Kinase binding	• Mitochondrial dysfunction*	
ALS/PD	N/A	• (Tissues: ALS – primary motor and sensory cortex, cervical spinal cord, lumbar spinal cord, blood; PD – substantia nigra, frontal gyrus, putamen, cerebellum, occipital cortex, blood) • n = 21 PD and 148 ALS genes (Note: these two genesets are from literature and reported pathways(COOPER-KNOCK *et al.* 2012) consistent between the two genesets are shown next) • RNA splicing* • Protein turnover*	N/A
AD/HD	N/A	N/A	• (Tissues: HD and AD – dorsolateral prefrontal cortex) • n = 3021 genes (in the shared DC network(NARAYANAN *et al.* 2014)) • Neurogenesis • Regulation of cell proliferation • Metabolic processes • n = 242 genes (in the 242-gene subnetwork(NARAYANAN *et al.* 2014)) • Neuron differentiation and neurogenesis* • Regulation of cellular metabolism* • Gap junction trafficking*

Unedited Version

			• Regulation of apoptosis* • Actin cytoskeleton complex* • Neurogenesis • Regulation of cell death • Response to wounding • Actin filament process

Unedited version

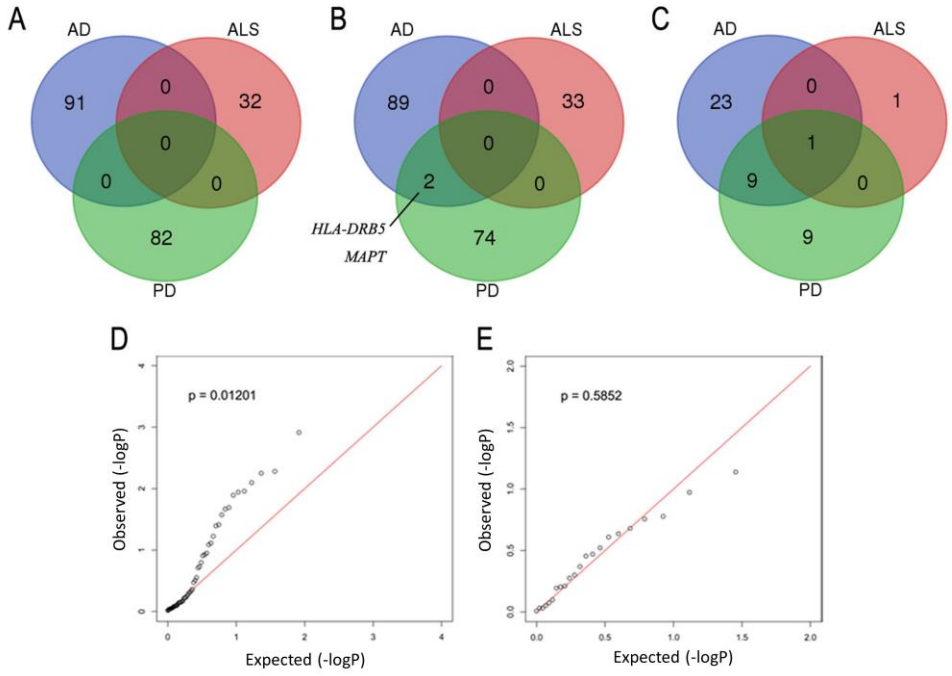

Figure 1. Overlap among GWAS genetic signals of AD, ALS and PD at SNP, candidate gene and pathway levels. Overlap among different diseases based on overlap among each disease's (A) GWAS-associated genome-wide significant ($p<5x10^{-8}$) SNPs, (B) candidate genes mapped to the genome-wide significant disease SNPs, and (C) pathways over-represented in the candidate disease genes. (D) QQ-plot showing that the significant PD SNPs ($p\leq5x10^{-8}$) collectively demonstrate more significant association with AD in AD GWAS (based on the full summary statistics from the IGAP stage1 study(LAMBERT *et al.* 2013)) compared to random expectation (p=0.01 based on the Kolmogorov–Smirnov or KS test). (E) QQ-plot showing that the significant ALS SNPs ($p\leq5x10^{-8}$) do not collectively demonstrate more significant association with AD in AD GWAS (based on the full summary statistics from the IGAP stage1 study(LAMBERT *et al.* 2013)) compared to random expectation (p = 0.59 based on KS test).

Unedited Ver

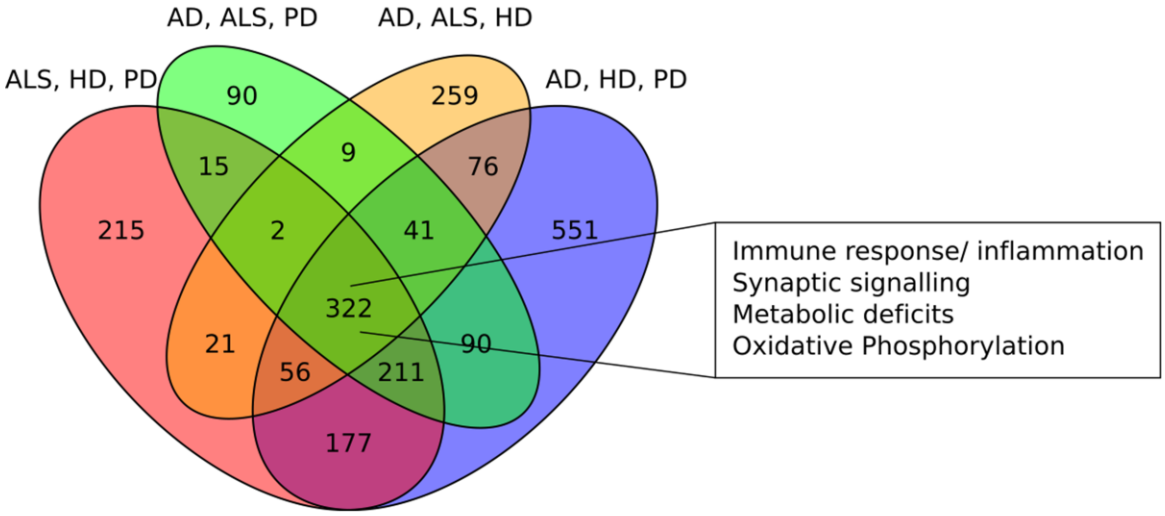

Figure 2. Overlap of differentially expressed genes in neurodegenerative diseases. Overlap of meta analyses of all possible combinations of three of the following diseases: AD, ALS, HD, and PD. 322 genes were consistently found across all four meta-analyses, and the enriched pathways of these conserved genes are shown. Note that these 322 conserved genes differ from the 243 consistent genes reported in the original study(Li *et al.* 2014) due to some differences in methodology - i.e., our approach of overlapping genes from the four meta analyses shown in this figure (as necessitated by the lack of ready availability of individual disease gene signatures from the original study) differs from the original study's approach of a single meta-analysis of all four diseases followed by validation in a replication cohort to derive the consistent genes.

Unedited Version

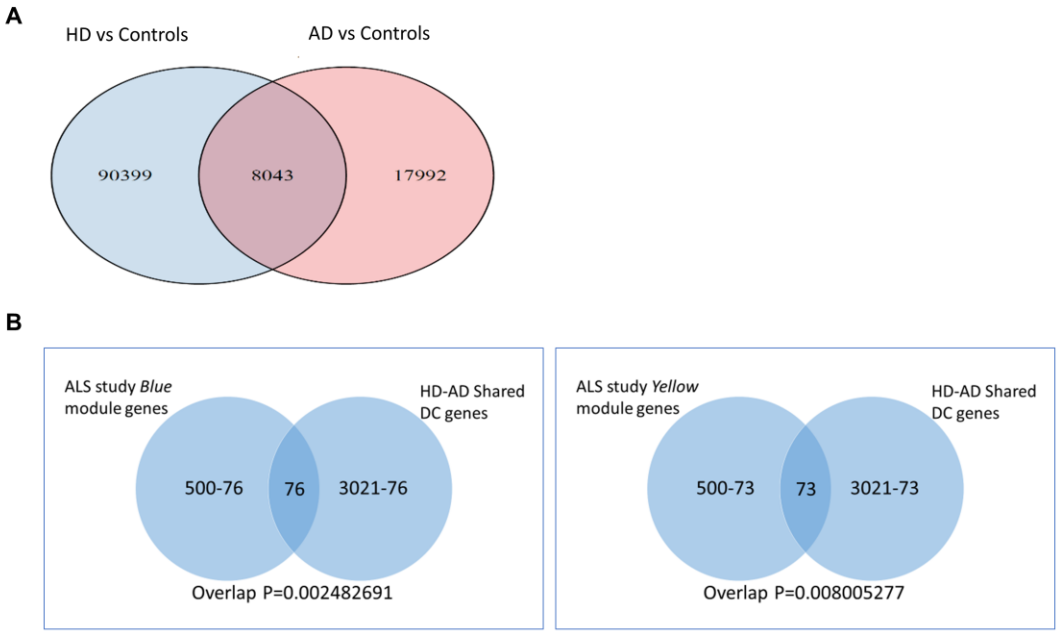

Figure 3. Shared network-level dysregulation in AD, HD and ALS. (A) The number of transcriptional network disruptions (DC or Differential Coexpression gene-gene relations) that were identified in our previous study(NARAYANAN *et al.* 2014) as AD-specific, HD-specific, or shared between both diseases are shown. The shared DC network comprised 8043 DC relations as shown (involving 3021 genes), and contained a majority of LOC (Loss of Correlation) relations, since the proportion of LOC among all DC relations in the AD, HD and shared DC networks were 33%, 19% and 51%, respectively. (B) Overlap of two coexpression network module genes identified from blood gene expression analysis of multiple ALS cohorts(SARIS *et al.* 2009) with the HD-AD shared DC network genes(NARAYANAN *et al.* 2014). The hypergeometric distribution based overlap p values compare the overlap of the module genes against the overlap rate of background genes (note that 1711 of the 15462 ALS study background genes overlapped with the set of 3021 HD-AD shared DC genes).

Unedited Ve