

Bioinformatics Approach to Identify Diseasome and Comorbidities Effect of Mitochondrial Dysfunctions on the Progression of Neurological Disorders

Md. Shahriare Satu¹, Koushik Chandra Howlader², Tajim Md. Niamat Ullah Akhund³, Fazlul Huq⁴, Julian M.W. Quinn⁵, and Mohammad Ali Moni^{4,5}

¹Dept. of CSE, Gono Bishwabidyalay, Dhaka, Bangladesh

²Dept. of CSTE, Noakhali Science and Technology University, Noakhali, Bangladesh

³Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh

⁴School of Biomedical Science, Faculty of Medicine and Health, The University of Sydney, Australia

⁵Bone Biology Division, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

Abstract Mitochondrial dysfunction can cause various neurological diseases. We therefore developed a quantitative framework to explore how mitochondrial dysfunction may influence the progression of Alzheimer's, Parkinson's, Huntington's and Lou Gehrig's diseases and cerebral palsy through analysis of genes showing altered expression in these conditions. We sought insights about the gene profiles of mitochondrial and associated neurological diseases by investigating gene-disease networks, KEGG pathways, gene ontologies and protein-protein interaction network. Gene disease networks were constructed to connect shared genes which are commonly found between the neurological diseases and Mitochondrial Dysfunction. We also generated KEGG pathways and gene ontologies to explore functional enrichment among them, and protein-protein interaction networks to identify the shared protein groups of these diseases. Finally, we verified our biomarkers using gold benchmark databases (e.g., OMIM and dbGaP) which identified effective reasons of it. Our network-based methodologies are useful to investigate disease mechanisms, predictions for comorbidities and identified distinct similarities among different neurological disorders for mitochondrial dysfunction.

Keywords

Mitochondrial Dysfunction; Neurological Diseases; Pathways; Ontology; Disease Association

Introduction

Mitochondria are organelles responsible for cellular energy metabolism, and mitochondrial dysfunction (MtD) underlie a heterogeneous group of chronic diseases [1], directly driving some pathological processes (e.g., in mitochondrial myopathies) and indirectly exacerbating others. Patients suffering from MtD may also lack proper diagnosis due to a lack of awareness about MtD and its consequences among clinicians [2]. MtD can result in insufficient ATP due to reduced mitochondrial numbers or defective oxidative phosphorylation, and can be caused by mutations in either mitochondrial and nuclear DNA. Mitochondria are increasingly viewed as having a central role in human health and disease, and their dysfunction may even be the root of some common neurological disorders [2]. Indeed, MtD has been implicated in the progression of Alzheimer's disease (AD) [3], Parkinson's disease (PD) [4], Huntington's Disease (HD) [5] and amyotrophic lateral sclerosis (ALS) [6] and cerebral palsy (CP).

A comorbidity refers to a co-incidence of distinct diseases in the same patient, that can result in disease interactions at the molecular level with serious repercussions for patient health and disease development [7, 8]. These associations can arise from causal relationships and shared risk factors among the diseases [9, 10]. Different diseases are associated with commonly dysregulated genes [11]. Studies of gene-disease relations helps us to identify significant interactions of molecular mechanism about disease [12]. Examples of comorbidity studies with bidirectional relationship happen between multiple variables include mental disorders, cancer, immune-related diseases and obesity. [13, 14, 15, 16, 17].

AD is the most common chronic neurodegenerative disease and is an irreversible, progressive disorder that can damage memory, analytical skills and the ability to perform elementary tasks [18]. The neuropathology of AD is characterized by cerebral accumulation and aggregation of amyloid- β ($A\beta$) and tau proteins [19] although their causal role is disputed. Cytochrome Oxidase (CO), a key enzyme of the mitochondrial electron transport chain can have reduced activity in AD patients and this may occur due to mitochondrial DNA (mtDNA) mutations [20]. PD is the second most common neurodegenerative disorder of aging that mostly affects the motor system. The clinical features of this disease are characterized by a loss of dopaminergic function which reduces motor function [21]. Significant evidence suggests that MtD can play a significant role in the pathogenesis of PD through inhibiting complex I of the electron transport chain. It is notable that several PD-associated genes participate in pathways regulating mitochondrial function, morphology, and dynamics [22]. HD is an inherited neurodegenerative disorder caused by gene mutations in huntingtin (Htt) which is characterised by progressive cognitive impairment, psychiatric disturbances and choreiform movements. Mutant Htt (mHtt) can be associated with MtD and energy metabolism defects that result in cell death [23]. ALS (also called Lou Gehrig's disease) is associated with a gradual loss of motor neurons of the brain-stem

and spinal cord. There is strong evidence that MtD is involved in ALS), but the underlying mechanisms linking MtD to motor neuron degeneration in ALS is uncertain. One possibility is that MtD give rise to SOD1 mutant proteins which damage motor neurons that causes ALS [24]. CP is an umbrella term for conditions that affect muscle tone, movement and motor skills in children [25]. Development of these neurological disorders may also involve a mtDNA mutation resulting in MtD [26].

The pathways that link MtD, its associated neurological diseases and biomarkers are essential to determine, in order to the coincidental occurrence of these diseases with MtD. We therefore initiated comprehensive KEGG pathways and gene ontology (GO) studies to explore these gene-disease associations, linking the diseases and genes from a biological perspective. To identify intersecting pathways affected by MtD which may influence neurons, we analyzed transcriptome evidence and investigated in details common gene expression profiles of AD, PD, HD, ALS, CP and MtD. From this analysis, we found possible common pathways arising from their common patterns of gene expression and studied these pathways using gold benchmark datasets including OMIM, dbGaP and protein protein interaction (PPI) data. This network based approach identified significant common pathways and pathway elements of significant potential interest in these neurological diseases.

Methods & Materials

A multi-stage analysis method is applied to analyze gene expression microarray data of various resouses of MtD and associated neurological diseases. We summarized our experimental approaches in Figure 1, where a systematic and quantitative approach was considered to assess human disease comorbidities employing existing microarray data. This approach identifies differentially expressed genes (DEGs) where functional enrichment studies was investigated enriched pathways, GO annotation terms, protein protein interactions, associated relations to identify putative pathways of common pathways and justified them with gold benchmark datasets.

Datasets employed in this study

To evaluate the molecular pathways involved in MtD on AD, PD, HD, ALS and CP at the molecular level, we first investigated gene expression microarray datasets. These microarray datasets were gathered from Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/geo/>) [27]. Six different datasets were selected with accession numbers such as GSE28146, GSE54536, GSE24250, GSE833, GSE31243 and GSE13887. The microarray AD dataset (GSE28146) is an affymetrix human genome U133 plus 2.0 array where laser capture microdissection is used by excluding significant white matter tracts. They also gather CA1 hippocampal gray matter from formalin-fixed, paraffin-embedded (FFPE) hippocampal parts of the 30 subjects of

their previous study. The PD dataset (GSE54536) is investigated 10 samples of PD patients where the whole transcriptome of peripheral blood of untreated people with stage 1 PD (HoehnYahr scale) was analyzed using Illumina HumanHT-12 V4.0 expression beadchip platform. The microarray HD (GSE24250) is also an affymetrix human genome U133A array where they indicate a dynamic biomarker of disease activity and treatment response needed to accelerate disease-modifying therapeutics for HD. The ALS (GSE833) is an affymetrix human full-length hugeneFL array where they compare gray matter associated genes sporadic and familial ALS patients compared with controls. The CP (GSE31243) is an affymetrix human genome U133A 2.0 array where 40 microarrays are provided into 4 groups to analyze the effects of cerebral palsy and differences between muscles. The MtD microarray dataset (GSE13887) is an affymetrix Human Genome U133 Plus 2.0 Array that show the activation of mammalian target of rapamycin (mTOR), sensor of the mitochondrial transmembrane potential can increase T cell in SLE.

Methodology

- The selection of GEO data was downloaded with matrices, platform information and converted expression set class for further analysis. Automatic selection of GEO data was not straightforward, because healthy and sick people were mixed together. So, experimental steps were helped to design model and factorize the classes of patients.

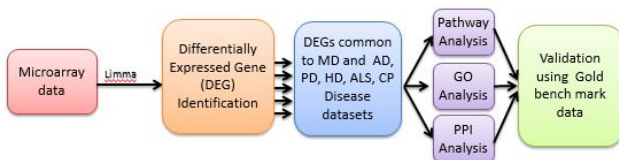


Figure 1. Flow Diagram of Network Based Approach to analyze MtD for Different Neurological Diseases

- We investigated gene expression profiles of MtD, AD, PD, HD, ALS and CP datasets using DNA microarray technologies with global transcriptome analyses. These datasets were detected DEGs with their respective pathology to compare disease tissues with normal one. This analysis was accomplished from original raw datasets and applied Limma R Bioconductor package for this microarray datasets. To overcome the problems of comparing microarray data in different systems and gene expression data were normalized and evaluated for each sample (disease or control) with the Z-score transformation (Z_{ij}). We used $Z_{ij} = \frac{g_{ij} - \text{mean}(g_i)}{SD(g_i)}$ for each disease gene expression matrix, where SD is the standard deviation, g_{ij} represents the expression value of gene i in sample j . This transformation was permitted direct comparisons of gene expression values across samples and diseases.

Data were transformed by \log_2 and student's unpaired t-test was employed between two conditions to find out DEGs in patients over normal samples by preferring significant genes. A p -value for the t-tests $< 5 * 10^{-2}$ and threshold at least 1 \log_2 fold change were selected. Besides, a two-way ANOVA with Bonferroni's *post hoc* test was applied to explore statistical significance between groups (< 0.01). Gene symbol and title of different genes were extracted from each disease. Null Gene symbol records were discarded from each disease. We also explored unique genes both over and under expressed genes. The most important up and down-regulated genes were selected between individual disease and MtD.

- A gene disease network (GDN) was built regarding the connection of genes and diseases where network nodes can be either diseases or genes by employing neighborhood-based benchmarking and topological methods. This kind of network can be represented as a bipartite graph whether MtD is the center of this network. Diseases are associated when they shares at least one unusual dysregulated gene. Let a special set of human diseases D and genes G , gene-disease relations seek where gene $g \in G$ is connected with disease $d \in D$. When G_i, G_j , the sets of notable up and down-dysregulated genes adhered with diseases i and j respectively, the number of shared dysregulated genes (n_{ij}^g) are associated with both diseases i and j is as follows:

$$n_{ij}^g = N(G_i \cap G_j) \quad (1)$$

Co-occurrence indicates shared genes in the GDN and common neighbours identified employing the Jaccard Coefficient method [28], where the edge prediction score of the node pair is:

$$E(i, j) = \frac{N(G_i \cap G_j)}{N(G_i \cup G_j)} \quad (2)$$

where E is the set of all edges. Our own R software packages "comoR" [29], and "POGO" [30] had been used to calculate novel estimators of the disease comorbidity associations.

- Gene Set Enrichment Analysis (GSEA) is a statistical methods which are used to classify genes into several groups, taking them into common biological functions, chromosomal location or regulation. Genes from microarray or NGS are investigated by their differential expression, selected as representative of several diseases, correlated to GO terms and finding possible pathways linked to phenotype changes. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) analysis are performed both up-regulated and down-regulated genes of each pair of diseases. To explore later insight into the functional enrichment, biological process and annotations of the molecular pathway of MtD that overlapped with AD, PD,

HD, ALS and CP, we performed GO analysis and KEGG pathway using the DAVID (Database for Annotation, Visualization and Integrated Discovery) (<https://david-d.ncifcrf.gov/>) [31] and Enrichr (<https://amp.pharm.mssm.edu/Enrichr/>) [32] bioinformatics resources [33].

- Protein Protein Interaction (PPI) network is also constructed from the STRING database (<https://string-db.org>) [34] for the neurological diseases. Proteins are represented by nodes and interaction between two proteins are represented by edges. We generated a PPI network that represented protein association among MtD with different neurological (AD, PD, HD, ALS and CP) diseases. Thus, we used Markov Cluster Algorithm (MCL) [35] to split them into different sub networks.
- Two gold bench mark verified datasets such as OMIM and dbGap were used to justify the principle how MtD was related with different neurological diseases in network based approach. Online Mendelian Inheritance in Man (OMIM) (<https://www.omim.org>) is a administrated database for retrieving the genes of all known diseases, relevant disorders, and genotype-phenotype relationships. Thus, database of Genotypes and Phenotype (dbGaP) (<https://www.ncbi.nlm.nih.gov/gap>) is considered to explore significant genes of different diseases with a particular disease.

Results

DEG analysis

We used Limma (Bioconductor packages) to analyze DEGs of all human microarray datasets to compare disease affected tissues (MtD, AD, PD, HD, ALS and CP). The patterns of gene expression of MtD patient tissues were investigated and validated using global transcription analysis. In each dataset, DEG were consequently identified and evaluated by R Bioconductor packages. A threshold of $\log_2 1$ (2-fold) was increased or decreased in transcript levels and genes with false discovery rate (FDR) below 0.05 were considered a significant requirement for a gene to accept as a DEG. We found 2737, 1532, 996, 2488, 551, 588 unfiltered DEG for MtD, AD, PD, HD, ALS and CP respectively. In MtD, 1154 and 1583 genes were significantly up-regulated and down-regulated in this work. We also accomplished cross comparative analysis to explore the usual significant genes between MtD and other neurological diseases and found that MtD shares 38, 30, 12, 14 and, 14 up-regulated genes and 88, 28, 197, 29 and 28 down-regulated genes with AD, PD, HD, ALS and CP. We constructed up-down disease association network (GDN) to find out common DEG among neurological diseases with MtD both for up-regulated (see Figure 2) and down-regulated (see Figure 3) genes. In these diagram, nodes are indicated genes or diseases and edges are specified the association between diseases and genes using Cytoscape V 3.6.1 [36, 37]. If one or more genes are

related between two distinct diseases, then two diseases are comorbid.

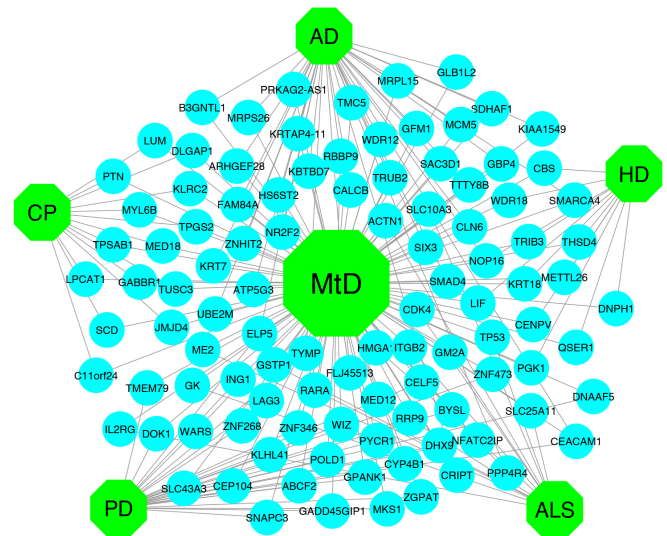


Figure 2. GDN with Up-Regulated (Increased Transcript Levels) DEGs

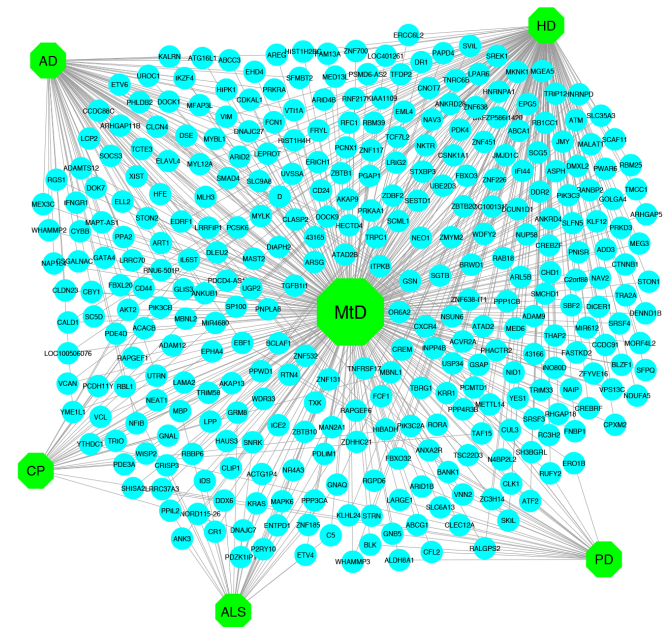


Figure 3. GDN of Down-Regulated (Decreased Transcript Levels) DEGs

Besides, some genes were found more than one common to MtD and neurological diseasedatasets. NOP16 is commonly up-regulated among MtD, AD, PD. Besides, TYMP, CEACAM1 and ELP5 is commonly up-regulated among MtD, AD, HD. On the other hand, FBXO32 and EBF1 are commonly down-regulated among MtD, AD, PD. CD24, FBXO32, RALGPS2, SKIL, CLK1, RTN4, N4BP2L2, ATF2, ZC3H14 and RUFY2 are commonly down-regulated among MtD, PD, HD. VIM, MLH3, SMAD4, WHAMMP3, DLEU2, DNAJC27, ARID2, MYL12A, LRRFIP1, UTRN, IL6ST, FBXO32, EDRF1, MYBL1 and LEPROT are commonly down-regulated among MtD, AD and HD. Fur-

thermore, UTRN, IL6ST and NFIB are commonly down-regulated among MtD, AD and CP. ZNF131 and GNAQ are commonly down-regulated among MtD, ALS and HD. LPP, MIR612, BCLAF1, MBNL2, MBNL1, IL6ST, UTRN, RORA and TAF15 are commonly down-regulated among MtD, HD and CP. DDX6 and DNAJC7 are commonly down-regulated among MtD, ALS and CP.

Functional enrichment of DEGs common to MtD and neurological diseases

We accomplished pathways and GOs on DEG sets (combination of up and down regulated genes) using DAVID v6.8 and Enrichr bioinformatics resources. In pathways analysis, KEGG data enrichment was only employed to manipulate MtD vs AD, MtD vs PD, MtD vs HD, MtD vs ALS and MtD vs CP. Integrating large scale, state of the art transcriptome and proteome analysis, a regulatory analysis was performed to obtain more insight about the molecular pathways related with these common genes and estimated links of affected pathways. KEGG pathway database (<https://www.genome.jp/kegg/pathway.html>) was used to explore pathways of DEGs where we pinpointed several overrepresented pathways among DEGs and classified them into functional groups. We observed that 5, 3, 5, 5 and 10 significant pathways, associated pathway genes and adjusted p-values are identified on AD, PD, HD, ALS and CP respectively using (<https://david-d.ncifcrf.gov/>) [31] and Enrichr (<https://amp.pharm.mssm.edu/Enrichr/>) (For exploring the pathways of PD) which are significantly connected with DEGs of MtD (see table1-A,B,C,D and E). Pathways esteemed common noteworthy enriched DEG sets (FDR < 0.05) were reduced by including only known relevance to the diseases concerned. On the other hand, we selected over characterized ontological groups among DEGs and also classified them into individual functional groups. 6, 12, 11, 11 and 10 significant GO groups, associated genes in the pathways and adjusted p values are explored for AD, PD, HD, ALS and CP respectively (see Table 2-A,B,C,D and E) where DEGs set are also common with MtD. Besides, we noticed a number of relevant significant pathway for example Insulin resistance was found as a common significant pathways among MtD, HD and CP.

Protein-protein interaction (PPI) Network Analysis

Malfunction of a protein complex causes multiple diseases. If they share one or more potentially related genes with each other through protein protein interaction network, then multiple diseases are potentially connected to each other. Thus, topological analysis of this network is the most important things to explore different associated protein of different diseases. Identified genes which involved in pathways and activities regular to MtD and different neurological diseases, we attempted to find evidence of existing sub-networks based on known PPI. Considering the enriched common disease genesets, we

built a PPI networks using web-based visualisation resource named STRING (see Figure 4). 413 proteins nodes are associated by 302 edges with PPI enrichment p-value 0.0486. 118, 34, 30 GO and 1 pathways are significantly enriched of Biological Process, Molecular Function, Cellular Component and KEGG Pathways respectively of this network. In this purpose, MCL clustering technique was performed to cluster proteins and many subnetwork contained proteins within one cluster. It indicated that PPI sub-network remained in our enriched genesets and clarified the presence of relevant functional pathways.

Comorbidity and Significant Pathway Analysis

Considering both gene expression data in MtD and gene disease association, GDN is constructed that can be explored gene disease association and comorbidity network. Multiple neurological disorders are connected to each other if they share at least one gene where nodes represent diseases or genes in which mutations have previously been related with both diseases. Each of these diseases HD (209 genes), AD (126 genes), PD (58 genes), ALS (43 genes) and CP (42 genes) are strongly associated with MtD since highest number of genes shared among them. We generated biologically relevant network projections which consisted of two disjoint sets of nodes where one set is corresponding to all known genetic disorders and other set is indicated our identified significant genes for MtD. Notably, 16 significant genes (SMAD4, NOP16, VIM, MLH3, SMAD4, DLEU2, DNAJC27, ARID2, MYL12A, LRRFIP1, UTRN, IL6ST, FBXO32, EDRF1, MYBL1, and LEPROT) are commonly dysregulated genes among AD, HD, and MtD, while 5 significant genes (TYMP, CEACAM1, ELP5, FBXO32, and EBF1) are widely dysregulated genes among PD, AD, and MtD. Then, 10 significant genes (CD24, FBXO32, RALGPS2, SKIL, CLK1, RTN4, N4BP2L2, ATF2, ZC3H14, and RUFY2) are commonly dysregulated genes among PD, HD, and MtD. Only 1 significant gene (RBL1) is commonly dysregulated genes among AD, ALS, and MtD. 3 significant genes (UTRN, IL6ST, and NFIB) are widely dysregulated genes among AD, CP, and MtD. Furthermore, 2 significant genes (ZNF700 and IFI44) are commonly dysregulated genes among HD, ALS, and MtD. 8 significant genes (LIF, AKAP9, BCLAF1, ANKRD28, GNAQ, CCDC91, GSAP and MEG3) have commonly dysregulated genes among HD, CP, and MtD and 2 significant genes (MBP and TXK) are widely dysregulated genes among ALS, CP, and MtD.

Validating Biomarkers by Gold Benchmark Databases

The OMIM and dbGaP databases were used to identify genes using single nucleotide polymorphism (SNP) that associated with different diseases (see Figure 5). It indicated that MtD associated genes are also responsible for various diseases. To find out significant neurological diseases, we found different diseases whose adjusted p-value were considered below or equal 0.05. Then, several diseases such as cancer, infectious diseases etc. were removed from this list because they were not concerned in

Table 1. KEGG pathway analyses which include significant pathways common to MtD and A) AD B) PD C) HD D) ALS E) CP. Pathways, genes of these pathways and adjusted p-values are given.

A. Common significant pathways of MtD and AD			
KEGG ID	Pathway	Genes in the pathway	Adj. p-value
hsa00532	Glycosaminoglycan biosynthesis - chondroitin sulfate /dermatan sulfate	CSGALNACT1;DSE	0.007
hsa04611	Platelet activation	LCP2;MYL12A;MYLK	0.041
hsa04670	Leukocyte transendothelial migration	CLDN23;CYBB;MYL12A	0.038
hsa04810	Regulation of actin cytoskeleton	GSN;DOCK1;MYL12A;MYLK	0.045
hsa04110	Cell cycle	SMAD4;RBL1;MCM5	0.043
B. Common significant pathways of MtD and PD			
KEGG ID	Pathway	Genes in the pathway	Adj. P-value
hsa04672	Intestinal immune network for IgA production	CXCR4;TNFRSF17	0.009
hsa04727	GABAergic synapse	SLC6A13;GNB5	0.029
hsa00240	Pyrimidine metabolism	POLD1;TYMP	0.040
C. Common significant pathways of MtD and HD			
KEGG ID	Pathway	Genes in the pathway	Adj. p-Value
hsa04550	Signaling pathways regulating pluripotency of stem cells	LIF, ACVR2A, IL6ST, SMAD4, SKIL, CTNNA1	0.011
hsa04520	Adherens junction	SMAD4, YES1, TCF7L2, CTNNA1	0.030
hsa03040	Spliceosome	SRSF3, SRSF4, TRA2A, HNRNPA1, RBM25	0.038
hsa04922	Glucagon signaling pathway	PPP4R3B, GNAQ, PRKAA1, ATF2	0.068
hsa04931	Insulin resistance	MGEA5, TRIB3, PRKAA1, PPP1CB	0.084
D. Common significant pathways of MtD and ALS			
KEGG ID	Pathway	Genes in the pathway	Adj. p-value
hsa05203	Viral carcinogenesis	KRAS, RBL1, TP53, ACTN1, CDK4	0.009
hsa05214	Glioma	KRAS, TP53, CDK4	0.028
hsa04720	Long-term potentiation	KRAS, GNAQ, PPP3CA	0.029
hsa05218	Melanoma	KRAS, TP53, CDK4	0.033
hsa05220	Chronic myeloid leukemia	KRAS, TP53, CDK4	0.034
E. Common significant pathways of MtD and CP			
KEGG ID	Pathway	Genes in the pathway	Adj. p-value
hsa04024	cAMP signaling pathway	PIK3CB, GABBR1, PDE3A, PDE4D, AKT2	0.003
hsa04510	Focal adhesion	LAMA2, PIK3CB, RAPGEF1, AKT2, VCL	0.003
hsa05146	Amoebiasis	LAMA2, GNAL, PIK3CB, VCL	0.004
hsa04910	Insulin signaling pathway	PIK3CB, ACACB, RAPGEF1, AKT2	0.008
hsa05211	Renal cell carcinoma	PIK3CB, RAPGEF1, AKT2	0.016
hsa05032	Morphine addiction	GABBR1, PDE3A, PDE4D	0.031
hsa04915	Estrogen signaling pathway	PIK3CB, GABBR1, AKT2	0.036
hsa05142	Chagas disease (American trypanosomiasis)	GNAL, PIK3CB, AKT2	0.039
hsa04931	Insulin resistance	PIK3CB, ACACB, AKT2	0.042
hsa05145	Toxoplasmosis	LAMA2, PIK3CB, AKT2	0.049

this study. After analyzing them, 3 neurological diseases were found in figure 5. Then, we constructed a GDN using Cytoscape and showed gene disease association of different diseases. It indicated that our analysis of finding significant genes of neurological diseases were also matched with existing records.

Discussion

The aim of this work is to find out the effectiveness of extracted information of bioinformatics analyses that can investigate the relationship of different neurological diseases and its comorbidities for MtD. We identified gene expression data from microarray experiments available in online and combined analysis of transcriptomics, genetics, pathways, GO data and PPI that have not been fulfilled by previous studies. So, It can fill up a significant

gap of our knowledge how MtD may affect AD, PD, HD, ALS and CP. A simple sequence of steps were applied to identify a range of possible comorbidities of neurological diseases. We investigated MtD by using selected pathologies and found experimental results which strengthen the scientific review of happening neurological diseases. Notably, we expect more and more microarray data that can be used to enrich the set of evidence without affecting previous outcomes. Therefore, It is suggested to take different datasets from various sources and cell types and integrated them to find out more robust evidences. The availability of suitable microarray datasets would be good to take not only classic class of case vs control, but also different conditions which comes from genetic variants to define the risk of having neurological diseases. In addition, the typology of data is important because all of the datasets have not contained similar standard. It is mean-

Table 2. GO analysis to identify Significant pathway common to MtD and A) AD B)PD C) HD D) ALS E) CP

A. Common significant GOs of MtD and AD			
GO Term	Pathways	Genes in the pathway	Adj. p-value
GO:0006366	transcription from RNA polymerase II promoter	GLIS3, EBF1, GATA4, SMAD4, SIX3, TGFB111, MYBL1, ETV6, NFIB	0.016
GO:0001932	regulation of protein phosphorylation	CCDC88C, SOCS3, ITPKB	0.019
GO:0060235	lens induction in camera-type eye	HIPK1, SIX3	0.025
GO:1904261	positive regulation of basement membrane assembly involved in embryonic body morphogenesis	CLASP2, PHLDB2	0.031
GO:0035726	common myeloid progenitor cell proliferation	ITPKB, CEACAM1	0.043
GO:0010470	regulation of gastrulation	CLASP2, PHLDB2	0.049
B. Common significant GOs of MtD and PD			
GO Term	Pathways	Genes in the pathway	Adj. p-value
GO:1902110	positive regulation of mitochondrial membrane permeability involved in apoptotic process	ATF2	0.027
GO:1902108	regulation of mitochondrial membrane permeability involved in apoptotic process	ATF2	0.024
GO:0000002	mitochondrial genome maintenance	TYMP	0.041
GO:1903608	protein localization to cytoplasmic stress granule	DHX9	0.024
GO:0010835	regulation of protein ADP-ribosylation	WARS	0.027
GO:0042987	amyloid precursor protein catabolic process	ABCG1	0.027
GO:0034374	low-density lipoprotein particle remodeling	ABCG1	0.035
GO:0042982	amyloid precursor protein metabolic process	ABCG1	0.041
GO:1901881	positive regulation of protein depolymerization	CFL2	0.038
GO:0060049	regulation of protein glycosylation	WARS	0.038
GO:0035269	protein O-linked mannosylation	LARGE1	0.035
GO:0006469	negative regulation of protein kinase activity	CEACAM1;WARS	0.043
C. Common significant GOs of MtD and HD			
GO Term	Pathway	Genes in the pathway	Adj. p-value
GO:0006397	mRNA processing	ZC3H14, SREK1, MBNL2, MBNL1, HNRNPA1, RBBP6, SRSF4, SCAF11, PAPD4, SFPQ, RBM39, WDR33, RBM25	6E-07
GO:0008380	RNA splicing	SRSF4, SCAF11, SFPQ, SREK1, MBNL2, RBM39, ZNF638, MBNL1, RBM25	4E-04
GO:0006357	regulation of transcription from RNA polymerase II promoter	ZMYM2, KLF12, ARID4B, ZNF131, MED13L, TCF7L2, JMY, ATF2, MED6, BLZF1, BRWD1, TFDP2, CHD1, LRRFIP1	1E-03
GO:0000209	protein polyubiquitination	CUL3, UBE2D3, RC3H2, RNF217, FBXO3, RBBP6, TRIP12	1E-02
GO:0036092	phosphatidylinositol-3-phosphate biosynthetic process	PIK3C2A, PIK3C3, INPP4B, ATM	2E-02
GO:0045892	negative regulation of transcription, DNA-templated	BCLAF1, SP100, TRIM33, CREBZF, SFPQ, SMAD4, TFDP2, MEG3, TRIB3, LRRFIP1, TCF7L2, CTNBN1	2E-02
GO:0044334	canonical Wnt signaling pathway involved in positive regulation of epithelial to mesenchymal transition	TCF7L2, CTNBN1	2E-02
GO:0051091	positive regulation of sequence-specific DNA binding transcription factor activity	SP100, DDR2, CTNBN1, ATF2, JMY	3E-02
GO:0031047	gene silencing by RNA	DICER1, RANBP2, CNOT7, TNRC6B, NUP58	3E-02
GO:0000398	mRNA splicing, via spliceosome	SRSF3, SRSF4, METTL14, TRA2A, HNRNPD, HNRNPA1, PPWD1	3E-02
GO:0043484	regulation of RNA splicing	MBNL2, CLK1, MBNL1	4E-02
D. Common significant GOs of MtD and ALS			
GO Term	Pathway	Genes in the pathway	Adj. p-value
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	CREM, RBL1, TP53, MED12, RARA, TXK, PPP3CA, NR4A3, HMGAI, ETV4	3E-04
GO:0045787	positive regulation of cell cycle	RARA, NR4A3, CDK4	3E-03
GO:0031100	organ regeneration	NR4A3, CDK4, GSTP1	5E-03
GO:0007411	axon guidance	KRAS, ANK3, NEO1, NR4A3	6E-03
GO:0035726	common myeloid progenitor cell proliferation	NR4A3, GSTP1	2E-02
GO:0009636	response to toxic substance	CDK4, GSTP1, MBP	2E-02
GO:2000774	positive regulation of cellular senescence	KRAS, HMGAI	2E-02
GO:0006468	protein phosphorylation	MAPK6, SNRK, RARA, TXK, CDK4	2E-02
GO:0050900	leukocyte migration	KRAS, PPIL2, ITGB2	3E-02
GO:0048384	retinoic acid receptor signaling pathway	CREM, RARA	4E-02
GO:0006367	transcription initiation from RNA polymerase II promoter	MED12, RARA, NR4A3	5E-02
E. Common significant GOs of MtD and CP			
GO Term	Pathway	Genes in the pathway	Adj. p-value
GO:0007155	cell adhesion	LAMA2, EPHA4, WISP2, LPE, ADAM12, VCL	0.003
GO:0071875	adrenergic receptor signaling pathway	AKAP13, PDE4D	0.004
GO:0030334	regulation of cell migration	LAMA2, AKT2, VCL	0.012
GO:0007185	transmembrane receptor protein tyrosine phosphatase signaling pathway	TRIO, PTN	0.013
GO:0006936	muscle contraction	MYL6B, UTRN, VCL	0.023
GO:0043065	positive regulation of apoptotic process	BCLAF1, TRIO, AKAP13, PTN	0.028
GO:0006198	cAMP catabolic process	PDE3A, PDE4D	0.033
GO:0061028	establishment of endothelial barrier	PDE4D, RAPGEF1	0.035
GO:0007165	signal transduction	GNAL, KLRC2, WISP2, PIK3CB, PDE4D, RAPGEF1, AKT2	0.040
GO:0010906	regulation of glucose metabolic process	RORA, ACACB	0.047

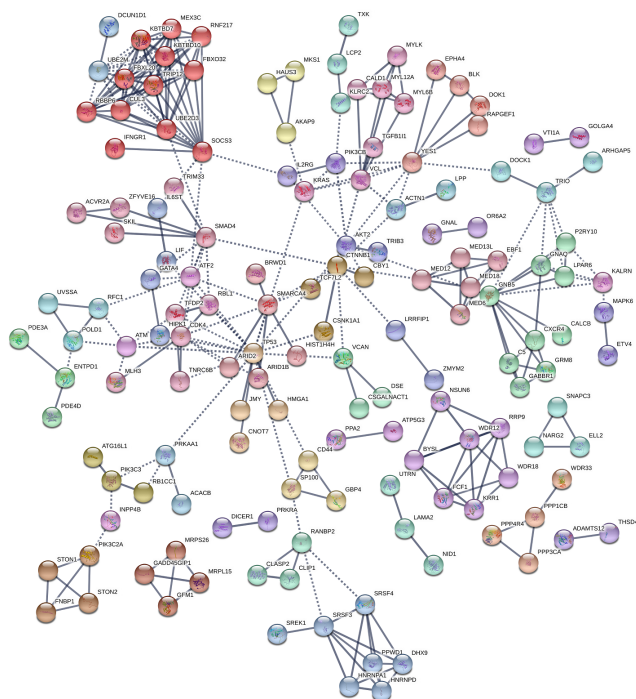


Figure 4. PPI network of DEGs with MCL Clustering

ingful that GSE series need more effort to prepare data for analysis. Before uploading microarray datasets, a personal quality control step is driven for semi automated analysis of them. Nevertheless, our approach is a good way to analyze, compare and integrate data and find a trade off from the realization of a standard protocols for MtD that causes various neurological diseases. Such analysis will be key element in the development of truly predictive medicine and provide insight to make novel hypothesis about diseases mechanisms of MtD. It will also predict the probability of developing disease comorbidity and relevant information about medication overlaps.

We applied a systematic and quantitative approach to compare how DEG transcripts in MtD that can be utilized to obtain insights into other human diseases. This work showed significant associations of neurological diseases with the corresponding MtD by substantial pathways. The integration of data obtained from different technologies which represents several ways to analyze MtD that causes distinct neurological diseases. Different neurological diseases are expressed potential gene-disease associations by combining analysis of genetics, regulatory patterns, pathways, GO data, and PPIs which have not been captured in previous studies. Our results show a combination of molecular and population-level data that can provide insights to happen neurological diseases for MtD. Furthermore, it will predict significant information about medication overlaps and the probability of developing disease comorbidity (where the occurrence of one disease may increase the susceptibility of other disease in a patient) using molecular biomarkers. To explore pathway and transcript profiles that can indicate disease associations and comorbid vulnerability for MtD. We analyzed

publicly available microarray data that can be used to investigate gene enrichment from signalling pathway and GO data. Besides, flexible, time consuming and semantic analysis-based approaches were used to facilitate this work and reduce operator bias. In transcript analyses, there were found an evidence about the processing of disease of PPI data. So, we can identify different pathways through inspection of cell proteins and their interactions. This investigation also represent a high potential of understanding the central mechanisms behind the neurological disease progression for MtD from the molecular and genetic aspects. This analyses will also be considered as a key element for predictive drugs development.

Table 3. Associated Diseases with Similar Genes

Term	Adj. p-value	Genes
ALS	2.02077E-09	AKAP13;GNAL;LAMA2;DLGAP1;ARSG
AD	2.6251E-09	ABCA1;SLC35A3;CR1;ARHGAP18;RORA
PD	3.30502E-08	AKAP13;BANK1;CREM;PHACTR2;ETV6

In summary, this work is significant from clinical bioinformatics perspective because this methodology is provided comorbidity and improved visualization aspects of MtD in neurological diseases. This process also harmonizes to the existing software that can use by physicians. The impact of bioinformatics in clinical investigation is still at the beginning. So, we hope that our approach is not only useful for the researches of neurological diseases but also other complex diseases.

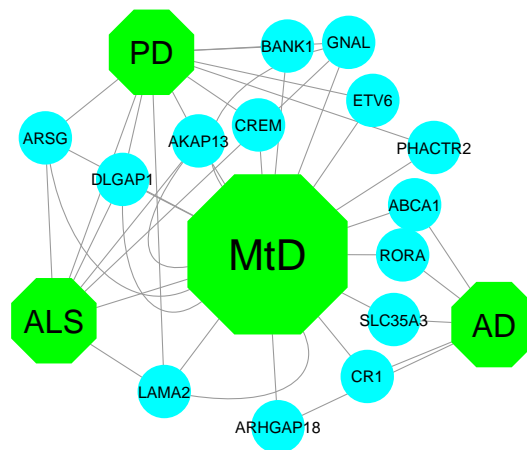


Figure 5. Gene identified by OMIM and dbGap databases using SNP related with MtD and diseases

Conclusion

This study defines the value of an integrated approach in disseminating individual neurological disease's relationships and expressed new opportunities for therapeutic applications. The multi-stage analysis methods were applied to analyze gene expression and regulatory data and DEGs were considered for functional enrichment studies to identify enriched pathways and GO terms of the biological processes (KEGG pathways, GO pathways and

PPI network), investigate disease mechanisms, comorbidities and regulated patterns that indicates how to impact MtD to the different neurological diseases. Genomic information-based personalized medicine is usually provided new fundamental insights about disease mechanisms. But, much genomic sequence information is not yet interpretable straightforwardly. Notably, genetic variant's effects on transcription control are slowly disclosed. In this situation, the reduction of sequencing costs can be enabled a revolution to analyze transcript that still provide huge health gains. So, our methods of DEGs analysis will need more development to analyze DEGs and will predict complimentary drugs for neurological diseases.

References

- [1] Richard H Haas, Sumit Parikh, Marni J Falk, Russell P Saneto, Nicole I Wolf, Niklas Darin, and Bruce H Cohen. Mitochondrial disease: a practical approach for primary care physicians. *Pediatrics*, 120(6):1326–1333, 2007.
- [2] Foundation for Mitochondrial Medicine related diseases. <http://mitochondrialdiseases.org/related-diseases/>. Accessed: 2018-07-20.
- [3] Renato X Santos, Sónia C Correia, Xinglong Wang, George Perry, Mark A Smith, Paula I Moreira, and Xiongwei Zhu. Alzheimer's disease: diverse aspects of mitochondrial malfunctioning. *International journal of clinical and experimental pathology*, 3(6):570, 2010.
- [4] Richard Kones. Parkinson's disease: mitochondrial molecular pathology, inflammation, statins, and therapeutic neuroprotective nutrition. *Nutrition in Clinical Practice*, 25(4):371–389, 2010.
- [5] M Flint Beal. Mitochondrial dysfunction in neurodegenerative diseases. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1366(1-2):211–223, 1998.
- [6] Philip C Wong, Carlos A Pardo, David R Borchelt, Michael K Lee, Neal G Copeland, Nancy A Jenkins, Sangram S Sisodia, Don W Cleveland, and Donald L Price. An adverse property of a familial als-linked sod1 mutation causes motor neuron disease characterized by vacuolar degeneration of mitochondria. *Neuron*, 14(6):1105–1116, 1995.
- [7] Mohammad Ali Moni and Pietro Liò. Network-based analysis of comorbidities risk during an infection: Sars and hiv case studies. *BMC bioinformatics*, 15(1):333, 2014.
- [8] César A Hidalgo, Nicholas Blumm, Albert-László Barabási, and Nicholas A Christakis. A dynamic network approach for the study of human phenotypes. *PLoS computational biology*, 5(4):e1000353, 2009.
- [9] Bin Tong and Chris Stevenson. *Comorbidity of cardiovascular disease, diabetes and chronic kidney disease in Australia*. Australian Institute of Health and Welfare, 2007.
- [10] Haoming Xu, Mohammad Ali Moni, and Pietro Liò. Network regularised cox regression and multiplex network models to predict disease comorbidities and survival of cancer. *Computational biology and chemistry*, 59:15–31, 2015.
- [11] Vincent Navratil, Benoit de Chasse, Chantal Rabourdin Combe, and Vincent Lotteau. When the human viral infectome and disease networks collide: towards a systems biology platform for the aetiology of human diseases. *BMC systems biology*, 5(1):13, 2011.
- [12] Aldo Segura-Cabrera, Carlos A García-Pérez, Xianwu Guo, and Mario A Rodríguez-Pérez. A viral-human interactome based on structural motif-domain interactions captures the human infectome. *PLoS one*, 8(8):e71526, 2013.
- [13] Klaas J Wardenaar and Peter de Jonge. Diagnostic heterogeneity in psychiatry: towards an empirical solution. *BMC medicine*, 11(1):201, 2013.
- [14] Lotte Holm Land, Susanne Oksbjerg Dalton, Trine Lembrecht Jørgensen, and Marianne Ewertz. Comorbidity and survival after early breast cancer. a review. *Critical reviews in oncology/hematology*, 81(2):196–205, 2012.
- [15] Alexandra Zhernakova, Cleo C Van Diemen, and Cisca Wijmenga. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nature Reviews Genetics*, 10(1):43, 2009.
- [16] Daphne P Guh, Wei Zhang, Nick Bansback, Zubin Amarsi, C Laird Birmingham, and Aslam H Anis. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC public health*, 9(1):88, 2009.
- [17] Mohammad Ali Moni, Haoming Xu, and Pietro Liò. Cytocom: a cytoscape app to visualize, query and analyse disease comorbidity networks. *Bioinformatics*, 31(6):969–971, 2014.
- [18] Alzheimer's disease fact sheet. <https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet>. Accessed: 2018-07-20.
- [19] Dennis J Selkoe. Alzheimer's disease results from the cerebral accumulation and cytotoxicity of amyloid β protein. *Journal of Alzheimer's disease*, 3(1):75–82, 2001.
- [20] Pierre Chagnon, Mark Gee, Mario Filion, Yves Robitaille, Majid Belouchi, and Denis Gauvreau. Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with alzheimer disease and controls in a french-canadian founder population. *American journal of medical genetics*, 85(1):20–30, 1999.
- [21] Anette Schrag, Laura Horsfall, Kate Walters, Alastair Noyce, and Irene Petersen. Prediagnostic presentations of parkinson's disease in primary care: a case-control study. *The Lancet Neurology*, 14(1):57–64, 2015.
- [22] Konstanze F Winklhofer and Christian Haass. Mitochondrial dysfunction in parkinson's disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1802(1):29–44, 2010.
- [23] Teresa Cunha-Oliveira, Ildete Luísa Ferreira, and A Cristina Rego. Consequences of mitochondrial dysfunction in huntington's disease and protection via phosphorylation pathways. In *Huntington's Disease-Core Concepts and Current Advances*. InTech, 2012.

- [24] Zhen Jiang, Wenzhang Wang, George Perry, Xiongwei Zhu, and Xinglong Wang. Mitochondrial dynamic abnormalities in amyotrophic lateral sclerosis. *Translational neurodegeneration*, 4(1):14, 2015.
- [25] Cerebral palsy. <https://kidshealth.org/en/parents/cerebral-palsy.html>. Accessed: 2018-07-20.
- [26] Alan Fryer, Richard Appleton, MG Sweeney, Lewis Rosenbloom, and AE Harding. Mitochondrial dna 8993 (narp) mutation presenting with a heterogeneous phenotype including cerebral palsy. *Archives of disease in childhood*, 71(5):419–422, 1994.
- [27] Home - geo datasets - ncbi. <https://www.ncbi.nlm.nih.gov/gds>. Accessed: 2018-06-12.
- [28] Mohammad Ali Moni and Pietro Lio. Genetic profiling and comorbidities of zika infection. *The Journal of infectious diseases*, 216(6):703–712, 2017.
- [29] Mohammad Ali Moni and Pietro Liò. comor: a software for disease comorbidity risk assessment. *Journal of clinical bioinformatics*, 4(1):8, 2014.
- [30] Mohammad Ali Moni and Pietro Liò. How to build personalized multi-omics comorbidity profiles. *Frontiers in cell and developmental biology*, 3:28, 2015.
- [31] Da Wei Huang, Brad T Sherman, and Richard A Lempicki. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature protocols*, 4(1):44, 2008.
- [32] Maxim V Kuleshov, Matthew R Jones, Andrew D Rouillard, Nicolas F Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L Jenkins, Kathleen M Jagodnik, Alexander Lachmann, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1):W90–W97, 2016.
- [33] GK Smyth, Matthew Ritchie, and N Thorne. Linear models for microarray data user’s guide. *Bioinformatics*, 20:3705–3706, 2011.
- [34] Damian Szklarczyk, John H Morris, Helen Cook, Michael Kuhn, Stefan Wyder, Milan Simonovic, Alberto Santos, Nadezhda T Doncheva, Alexander Roth, Peer Bork, et al. The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, page gkw937, 2016.
- [35] Anton J Enright, Stijn Van Dongen, and Christos A Ouzounis. An efficient algorithm for large-scale detection of protein families. *Nucleic acids research*, 30(7):1575–1584, 2002.
- [36] Michael Kohl, Sebastian Wiese, and Bettina Warscheid. Cytoscape: software for visualization and analysis of biological networks. In *Data Mining in Proteomics*, pages 291–303. Springer, 2011.
- [37] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003.