



A computational approach to identify blood cell-expressed Parkinson's disease biomarkers that are coordinately expressed in brain tissue



Mohammad Ali Moni^{a,i,*}, Humayan Kabir Rana^b, M. Babul Islam^c, Mohammad Boshir Ahmed^d, Haoming Xu^{e,f}, Md Al Mehedi Hasan^g, Yiming Lei^h, Julian M.W. Quinn^{i,**}

^a School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Australia

^b Department of Computer Science and Engineering, Green University of Bangladesh, Bangladesh

^c Department of Electrical and Electronic Engineering, University of Rajshahi, Bangladesh

^d School of Materials Science and Engineering, Bio-Electronics Materials Laboratory/ Gwangju Institute of Science and Technology, 261 Cheomdan-gwagiro (Oryong-dong), Buk-gu, Gwangju, 500-712, Republic of Korea

^e Department of Biomedical Engineering, Duke University, Durham, NC, 27708, USA

^f Chengdu Institute of Public Administration, Sichuan, 610110, China

^g Department of Computer Science and Engg, Rajshahi University of Engg. and Technology, Bangladesh

^h School of Electronics Engineering and Computer Sciences, Peking University, Beijing, China

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

ARTICLE INFO

Keywords:

Parkinson's disease
Brain
Blood
cis-eQTL
GWAS

ABSTRACT

Identification of genes whose regulation of expression is functionally similar in both brain tissue and blood cells could in principle enable monitoring of significant neurological traits and disorders by analysis of blood samples. We thus employed transcriptional analysis of pathologically affected tissues, using agnostic approaches to identify overlapping gene functions and integrating this transcriptomic information with expression quantitative trait loci (eQTL) data. Here, we estimate the correlation of gene expression in the top-associated cis-eQTLs of brain tissue and blood cells in Parkinson's Disease (PD).

We introduced quantitative frameworks to reveal the complex relationship of various biasing genetic factors in PD, a neurodegenerative disease. We examined gene expression microarray and RNA-Seq datasets from human brain and blood tissues from PD-affected and control individuals. Differentially expressed genes (DEG) were identified for both brain and blood cells to determine common DEG overlaps. Based on neighborhood-based benchmarking and multilayer network topology approaches we then developed genetic associations of factors with PD.

Overlapping DEG sets underwent gene enrichment using pathway analysis and gene ontology methods, which identified candidate common genes and pathways. We identified 12 significantly dysregulated genes shared by brain and blood cells, which were validated using dbGaP (gene SNP-disease linkage) database for gold-standard benchmarking of their significance in disease processes. Ontological and pathway analyses identified significant gene ontology and molecular pathways that indicate PD progression.

In sum, we found possible novel links between pathological processes in brain tissue and blood cells by examining cell pathway commonalities, corroborating these associations using well validated datasets. This demonstrates that for brain-related pathologies combining gene expression analysis and blood cell cis-eQTL is a potentially powerful analytical approach. Thus, our methodologies facilitate data-driven approaches that can advance knowledge of disease mechanisms and may, with clinical validation, enable prediction of neurological dysfunction using blood cell transcript profiling.

* Corresponding author. School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, NSW 2006, Australia.

** Corresponding author.

E-mail addresses: mohammad.moni@sydney.edu.au (M.A. Moni), humayan.pustcse@gmail.com (H.K. Rana), babul.apee@ru.ac.bd (M.B. Islam), Mohammad.Ahmed@gist.ac.kr (M.B. Ahmed), hmingxu@gmail.com (H. Xu), j.quinn@garvan.org.au (J.M.W. Quinn).

<https://doi.org/10.1016/j.combiomed.2019.103385>

Received 18 April 2019; Received in revised form 6 August 2019; Accepted 7 August 2019

Available online 09 August 2019

0010-4825/ © 2019 Elsevier Ltd. All rights reserved.

1. Introduction

The difficulty of early diagnosis and discrimination of neurodegenerative diseases such as Parkinson's Disease (PD), Alzheimer's (AD), Huntington's (HD) and motor neuron diseases (MND) mean that identifying robust biomarkers for these conditions in blood samples is a long held goal [1–6]. While gene expression studies of peripheral blood mononuclear cells cannot directly study the diseased central nervous tissues at issue, it is an approach with the potential to characterize the influences of systemic factors that similarly affect both brain and blood cells. These influences include genetic factors, notably protein coding-gene mutations and single nucleotide polymorphisms (SNPs), as well as other systemic diseases (i.e., comorbidities) that are suffered by neurodegenerative disease patients. The latter can be an important consideration since many comorbidities are known risk factors for neurodegenerative diseases [7–9]. However, since most genes are regulated differently in different tissues, to obtain useful biomarkers among blood cell transcripts we need to identify genes that are similarly regulated in brain and blood cells. Achieving this could enable blood cell transcript profiles to become a window for viewing some of the pathological changes affecting the brain.

Genome-wide association studies (GWAS) have discovered thousands of genetic variants associated with complex traits and pathological conditions, including neurodegenerative diseases [10,11]. With advances in microarray and RNA-Seq technologies, genome-wide sequencing together with tissue gene expression data from relatively large numbers of samples have been generated as part of attempts to identify genetic variants that affect transcript abundance [12,13], i.e., expression quantitative trait loci (eQTLs). The Genotype-Tissue Expression (GTEx) project [14] has provided a large and increasingly comprehensive resource of data that enable investigation of the genetic causes of gene expression variation across a broad range of tissues and cell types, particularly including blood cells and brain tissue. Molecular associations, evident from differential gene expression patterns, protein-protein interactions (PPIs), gene ontologies and common metabolic pathways can mediate the effects risk factors that influence or drive the development of a disease [15–17]. A risk factor (e.g., a pathogen or a comorbidity) and a disease may reveal a mechanistic link if both cause altered expression in a common set of genes [18–21]. In addition, from a proteomics and signaling pathways perspective, such links may also be found through the demonstrations of common biological modules such as PPIs, gene ontologies or molecular pathways [22–25]. We thus used a network based analyses to identify the influence of genetic factors and disorders in PD progression by utilizing the gene expression profiling, PPI sub-network, gene ontologies and molecular pathways. An extensive study regarding phylogenetic and pathway analysis was also conducted to reveal the genetic associations of the PD.

In this study, we have investigated PD, a progressively developing degenerative disorder that mainly damages the motor system of the central nervous system [26] suffered by over 10 million people worldwide [27]. PD damages dopaminergic neurons in the substantia nigra pars compacta and forms Lewy bodies, neuronal cell soma inclusions containing α -synuclein [28,29]. Subtle early symptoms seen in PD affected individuals comprise shaking, rigidity, slowness of movement and complications in mobility. PD patients suffer from difficulties in walking, talking or even completing simple daily activities, while sensory, sleep and emotional problems may also be evident, and can lead to development of dementia [30,31]. Currently the main causes or risk factors of PD remain poorly understood [32,33]. For example, it unclear why the risk of PD development is affected by gender or by physical exercise [34]. To gain mechanistic insight into PD (and potentially other central nervous system disorders) we investigated PD-associated transcripts (from RNA-Seq and microarray data) that are common to brain and blood cells, then used human-expressed quantitative trait loci (eQTL) data to identify genes that are expressed under similar expression control in both cell types. Further screens and

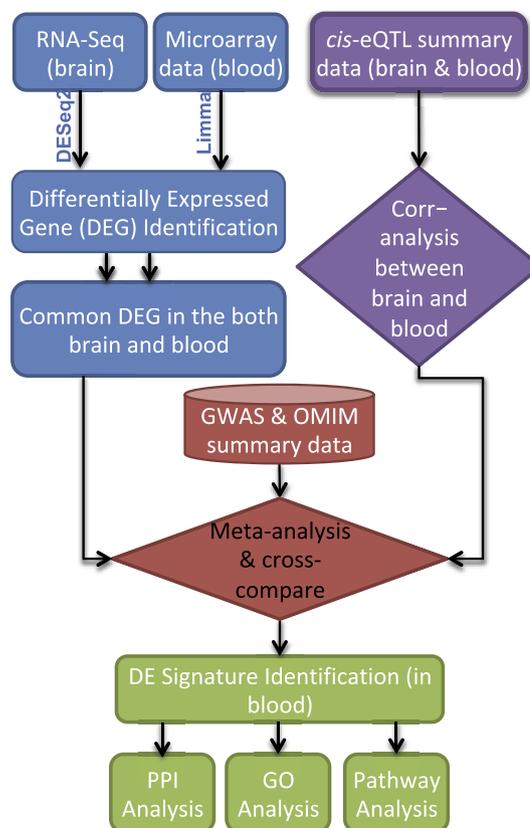


Fig. 1. Flowchart of the pipeline that is used for the early detection of the neurological dysfunction using blood cell transcripts.

filtering methods that employ human genetics and transcriptomics databases (including microarray and RNAseq data), as well as curated gold standard benchmark databases for PD was used to reveal potential brain biomarker genes and cell pathways which behave similarly in blood. Our overall approach aims to find overlapping pathways of potential clinical utility, but with the possibility that we can also identify important new pathways relevance to many neurological diseases by examining blood cell transcripts.

2. Materials and methods

2.1. Overview of analytical approach

We present here, summarised in Fig. 1, a systematic and quantitative approach to using blood cell information to predict neurological disorder at the early stages using available sources of mRNA expression, RNA-Seq, Genome Wide Association Studies (GWAS) and expression of quantitative trait loci (eQTL) datasets. This approach employs gene expression analyses, signalling pathway information, Gene Ontology (GO) data, disease-gene associations and protein-protein interaction data to identify putative components of common pathways for the neurological dysfunction evident in brain and blood cells.

2.2. Datasets employed in this study

To investigate the molecular pathways involved in PD, we first employed global transcriptome analyses (RNAseq datasets) as well as gene expression microarray datasets related to the blood and brain cells. In this study, we collected raw data from the Gene Expression Omnibus of the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/geo/>). We selected 2 different large human gene expression datasets for our study with accession numbers

GSE68719 and **GSE22491**. GSE68719 is an RNAseq dataset from a study of PD using brain cells from healthy and PD individuals. GSE22491 is an Affymetrix RNA array dataset from a study of normal and PD individuals whole blood cells. We have also used GWAS catalogues is the NHGRI-EBI Catalog of published genome-wide association studies (<https://www.ebi.ac.uk/gwas/>) and eQTL data (both blood and brain) from the GTExPortal which is a database for the Genetic Association data. (<https://gtexportal.org/home/>)

To get further insight into the molecular pathways of PD that overlap between brain and blood cells, we performed pathway and gene ontology analysis using the DAVID bioinformatics resources (<https://david.ncifcrf.gov/>) and KEGG pathways database [35]. We also generated a protein-protein interaction (PPI) network for each disease-pair datasets, using data from the STRING database string-db.org [36]. Furthermore, we also incorporated a gold benchmark verified Gene-Trait associations dataset of the dbGaP (www.ncbi.nlm.nih.gov/gap) in our study for validating the proof of principle of our network based approach.

2.3. Analysis methods

Using RNAseq and RNA microarray technologies for global transcriptome analyses, we compared the gene expression profiles of PD with that in brain and blood cells datasets. All these datasets were generated by comparing diseased tissue with normal to identify differentially expressed genes (DEG) associated with their respective pathology PD. The analysis was performed from the original raw datasets and employed DESeq2 [37] and Limma [38] R Bioconductor packages for the RNAseq and microarray data respectively. To avoid the problems of comparing mRNA expression data of different platforms and experimental systems, we normalized and calibrated the gene expression data in each sample (disease state or control) using the Z-score transformation (Z_{ij}) for each disease gene expression matrix using:

$$Z_{ij} = \frac{g_{ij} - \text{mean}(g_i)}{SD(g_i)} \quad (1)$$

where SD is the standard deviation, g_{ij} represents the expression value of gene i in sample j . This transformation allows direct comparisons of gene expression values across samples and diseases. We applied a Studentised t -test statistic between two conditions. Data were \log_2 -transformed and Student's unpaired t -test performed to identify genes that were differentially expressed in patients over normal samples and significant genes were selected. A threshold of at least $1 \log_2$ fold change and a p -value for the t -tests of $<5 \times 10^{-2}$ were chosen. In addition, a one-way ANOVA with Bonferroni's *post hoc* test was used to establish statistical significance between groups (<0.01).

For gene-disease associations, we applied neighborhood-based benchmarking and topological methods. Given a particular set of human diseases D and a set of human genes G , gene-disease associations

attempt to find whether gene $g \in G$ is associated with disease $d \in D$. If G_i and G_j , the sets of significant up- and down-dysregulated genes associated with diseases i and j respectively, then the number of shared dysregulated genes (n_{ij}^g) associated with both diseases i and j is as follows:

$$n_{ij}^g = N(G_i \cap G_j) \quad (2)$$

The common neighbours identified by an approach based on the Jaccard Coefficient method [6,39–41], where the edge prediction score for the node pair is:

$$E(i, j) = \frac{N(G_i \cap G_j)}{N(G_i \cup G_j)} \quad (3)$$

where E is the set of all edges. We have used our own R software packages “comoR” [42] and “POGO” [43] to compute novel estimators of the disease comorbidity associations.

2.4. Effects of eQTL between blood and brain

Let \bar{x} is the association of eQTL effect between blood and brain for a gene. We can measure the effect of eQTL between blood and brain \bar{x} as

$$\bar{x} = x + \varepsilon \quad (4)$$

where x is the absolute effect between blood and brain and ε is the estimated error. We assume that x and ε are random variables (i.e., $x \sim N(0, \text{var}(x))$ and $\varepsilon \sim N(0, \text{var}(\varepsilon))$) over the genes between blood and brain (see details in Ref. [14]).

3. Results

3.1. DEG analysis of datasets

DEG analysis of human RNAseq and microarray datasets, comparing disease affected tissues, was performed using DESeq2 and Limma (Bioconductor packages). We used publicly available RNAseq and microarray data for brain and blood cells in PD affected individuals and controls. Such studies compare affected vs unaffected individuals to reveal differentially expressed genes (DEG) for each disease. Genes with false discovery rate (FDR) below 0.05 and a threshold of $2 \log$ fold increase or decrease in transcript levels was required for a gene to be accepted as a DEG. The numbers of unfiltered DEG identified were 487 in brain and 1083 in blood datasets.

We also performed cross comparative analysis to find the common significant genes between brain and blood cells. DEG common in brain and blood cells of PD affected people were identified and are summarised graphically in Figs. 2 and 3. We observed that the number of common positive significant genes between brain and blood cells is 7, and their log fold changes and negative log p-values are shown in Fig. 2. Similarly, the number of negative dysregulated significant genes between brain and blood cells is 7, and their log fold changes and negative

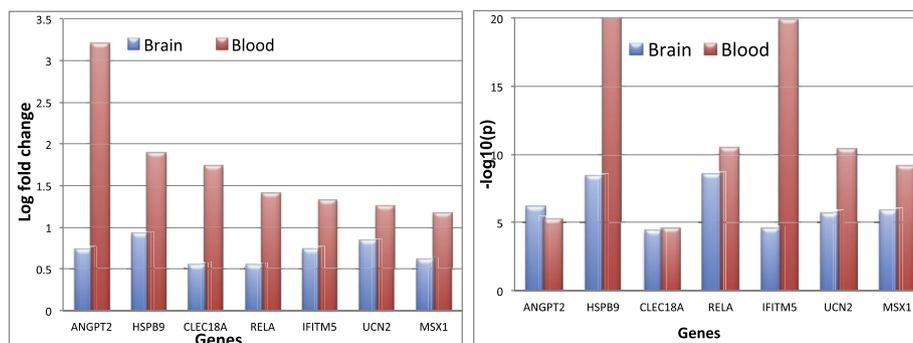


Fig. 2. Identification of upregulated DEG that is observed in both brain and blood. A) log fold changes of the common upregulated significant genes in brain and blood and B) Negative log of p-value of the upregulated significant genes common to brain and blood.

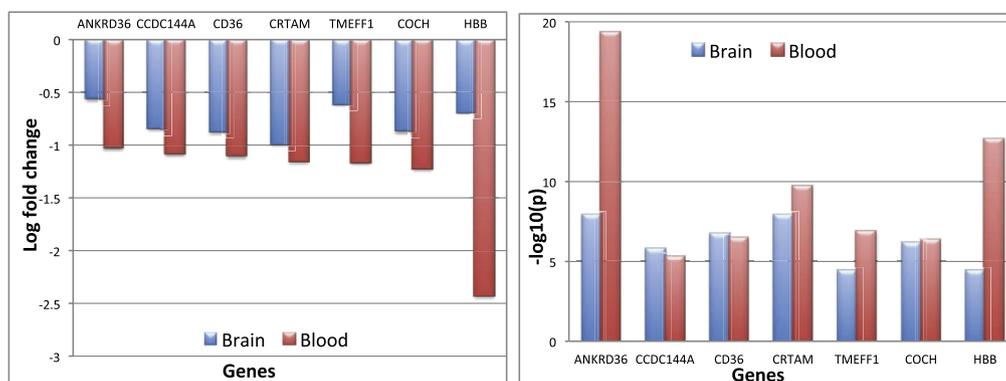


Fig. 3. Identification downregulated DEG in both brain and blood. A) log fold changes of the common downregulated significant genes in brain and blood and B) Negative log of p-values of downregulated significant genes common to brain and blood.

logarithmic p-values are shown in Fig. 3.

3.2. Identifying genes expressed in blood cells that mirror those expressed in brain

eQTL databases link gene SNPs to gene expression. Few such databases have been produced, but there are databases for blood and brain cells; we used these from the GTEx database to find genes with similar genetic control of expression in the two tissues using meta-analysis approaches. We have identified 673 such blood-brain co-expressed genes (BBCG) using the correlation and meta-analysis approach as explained in the method section.

3.3. Identifying genes expressed in blood cells that have brain expression associated in PD development

We selected genes whose expression or gene sequence variants (e.g., SNPs) reveal functional association with PD. This utilises curated gold-benchmark databases OMIM and GWAS catalogues. Using SNP and gene expression datasets from both brain and blood cells we identified BBCG among PD-associated genes. Similarly, using SNP and common cis-eQTL in brain and blood cells we also identified BBCG among PD-associated genes. Thus we identified 12 significant genes, C10orf32, CCDC82, COL5A2, COQ7, GPNMB, HSD17B1, KANSL1, NCKIPSD, PM20D1, SP1, FRRS1L and IL1R2 that are commonly dysregulated between blood and brain for the PD. Disease processes will alter their expression but systemic disease factors may similarly affect brain and blood cells. We used these potential PD biomarkers to identify pathways of regulation that are active in PD patients.

3.4. Identifying pathways in the blood cells that mirror in brain

We performed pathway and gene ontology analysis on DEG sets using DAVID bioinformatics resources. For pathways we used KEGG data enrichment was determined for the identified potential signature genes for PD progression in blood cells. To combine large scale, state of the art transcriptome and proteome analyses, we then performed a regulatory analysis in order to gain further insight into the molecular pathways associated with these common genes as well as predicted links to the affected pathways. DEG and pathways were analysed using KEGG pathway database (<http://www.genome.j/kegg/pathway.html>) and functional annotation analysis tool DAVID (<http://niaid.abcc.ncifcrf.gov>) to identify overrepresented pathway groups amongst DEG sets and to group them into functional categories. Pathways deemed significantly enriched in the common DEG sets (FDR < 0.05) were reduced by manual curation to include only those with known relevance to the diseases concerned. These data are summarised in Table 1. We observed a number of relevant and significant pathways

including TGF- β and MAPK signaling pathways.

To get further insight into the identified pathways, enriched common gene sets were processed by gene ontology methods using EnrichR (<http://amp.pharm.mssm.edu/Enrichr/>) which identifies related biological processes and groups them into functional categories. The list of processes was also curated for those with known involvement with the diseases of interest. The cell processes and genes identified are summarised in Table 2. We observed a number of significant pathways that notably included regulation of neuron death and negative regulation of neuron death.

3.5. Protein-protein interaction (PPI) analysis to identify functional sub-networks

Dysregulation in a protein subnetwork may yield dysfunctional multiple protein sub-networks. Several diseases may be caused or influenced by the malfunction of a particular protein complex [44]. Thus, two or more genes are potentially related to each other through their common association in a protein-protein interaction network. Having identified genes involved in pathways and processes common to brain and blood, we sought evidence for existing sub-networks based on known PPI. Using the enriched common disease genesets, we constructed putative PPI networks using web-based visualisation resource STRING [45]. Clustering of genes was also performed by the Markov cluster algorithm (MCL) and it was notable that many of the PPI network contained genes within one cluster, indicated in red in Fig. 4. This data provides evidence that PPI sub-network exist in our enriched genesets, and confirm the presence of relevant functional pathways.

4. Discussion

In this study, we investigated transcriptomic evidence for pathways in brain and blood cells that are similarly affected by PD. Employing global transcriptome analyses, we investigated in detail the common gene expression profiles of PD evident in both blood and brain cells identifying genes whose altered expression in the blood cells mirrors their altered expression in the brains cells, i.e., genes that, in the context of PD, might be used to identify altered gene expression in the brain by monitoring the blood. We sought these genes by considering eQTL databases which identify genes with similar expression control (identified where gene polymorphisms or sequence variants occur) in blood and brain cells. This in itself is not a new concept, as eQTL commonalities between blood and brain cells have been noted previously [46–48], although it is argued for some complex traits these common genes may have limited predictive value [49–51]. Nevertheless, eQTL databases catalogue the relationship between a polymorphism and its effects on transcript levels in the appropriate tissues (here, blood and brain) that can also indicate whether or not the

Table 1

KEGG pathway analyses to identify significant pathways for the identified potential biomarker for PD that revealed among genes expressed in common by brain and blood cells. Pathway genes and pathway adjusted p-values are indicated.

KEGG Term	Pathways	Adjusted P-value	Genes
hsa05202	Transcriptional misregulation in cancer	5.01E-03	SP1; IL1R2
hsa00130	Ubiquinone and other terpenoid-quinone biosynthesis	6.58E-03	COQ7
hsa04913	Ovarian steroidogenesis	2.96E-02	HSD17B1
hsa00140	Steroid hormone biosynthesis	3.43E-02	HSD17B1
hsa04350	TGF-beta signaling pathway	4.93E-02	SP1
hsa04010	MAPK signaling pathway	4.28E-02	IL1R2
hsa04060	Cytokine-cytokine receptor interaction	4.79E-02	IL1R2
hsa01100	Metabolic pathways	4.79E-02	HSD17B1; COQ7

transcripts are under similar control in different tissues. This knowledge has recently been used to identify possible gene regulation pathways [52,53]. Here, we have identified genes whose expression in the blood does indeed mirror that happening in the brain, although it should be noted that it is not possible to access blood and brain samples from the same individuals post mortem. Where blood cell and brain cell transcript mirroring is evident it is likely to arise from globally acting influences of the PD-related damage to central tissues, although what those influences might be need further study. It should be noted that any clinical utility of these genes that show such blood cell/brain coordinated expression would require proper clinical evaluation. Our focus here was to these gene data to identify common pathways that arise from this common patterns of gene expression; we then compared these with pathways seen in validated datasets including dbGaP (see Table 3) and from protein-protein interaction (PPI) data. These databases are usually referred to as a “gold” benchmark data since they are validated with published experimental biomedical data. This network-based approach identified significant common pathways influencing both brain and blood. Moreover, we used a novel computation-based approach that is based on our previous studies that developed analytical approaches in tissues exposed to environmental stimuli [9] to identify among the common gene expression pathways that are profoundly affected by the disease processes themselves as well as by predisposing genetic and environmental factors.

In this study, we have identified potential blood cell/brain coordinately expressed genes for PD-related processes that result in detectable changes to transcript levels in blood cells, i.e., we examined how DEG correlate with expression of PD-associated genes to find key genes with expression dysregulated in both brain and blood cells. These

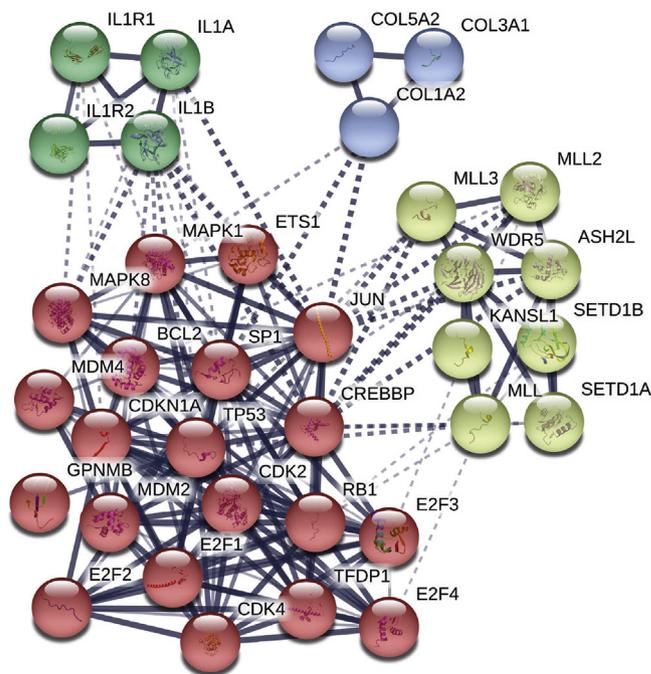


Fig. 4. Protein-protein interaction (PPI) network of the PD. These include significant pathways common to blood and brain cells as indicated. Genes were identified by STRING software tools. Colour indicates MCL analysis clusters of proteins. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2

Gene ontology identification of biological processes common to brain and blood cells in PD. KEGG pathway-enriched gene sets were employed in gene ontology (GO) studies to identify potential common processes. GO terms were curated to identify those relevant to brain and blood cells. Pathways genes and pathway adjusted p-values are indicated.

GO Term	Pathways	Adjusted P-value	Genes
GO:1901214	regulation of neuron death	8.47E-04	PM20D1; GPNMB
GO:0060548	negative regulation of cell death	1.07E-03	PM20D1; GPNMB
GO:1901215	negative regulation of neuron death	1.12E-03	PM20D1; GPNMB
GO:0044270	cellular nitrogen compound catabolic process	4.19E-03	PM20D1
GO:0090596	sensory organ morphogenesis	4.79E-03	COL5A2
GO:0045765	regulation of angiogenesis	4.90E-03	GPNMB; S PI
GO:0042773	ATP synthesis coupled electron transport	5.99E-03	COQ7
GO:0006743	ubiquinone metabolic process	8.97E-03	COQ7
GO:0008210	estrogen metabolic process	8.97E-03	HSD17B1
GO:0043981	histone H4-K5 acetylation	9.56E-03	KANSL1
GO:0034103	regulation of tissue remodeling	1.07E-02	GPNMB
GO:0006607	NLS-bearing protein import into nucleus	1.37E-02	NCKIPSD
GO:1901342	regulation of vasculature development	1.73E-02	GPNMB
GO:1900449	regulation of glutamate receptor signaling pathway	1.96E-02	FRRS1L
GO:0090181	regulation of cholesterol metabolic process	2.49E-02	SP1
GO:0050810	regulation of steroid biosynthetic process	2.67E-02	SP1
GO:0071822	protein complex subunit organization	2.73E-02	COL5A2
GO:0043603	cellular amide metabolic process	3.72E-02	PM20D1
GO:0008610	lipid biosynthetic process	4.29E-02	HSD17B1

Table 3

Genes significant in PD that may be potential markers in blood cells for PD. These genes are also significant for in dbGaP database using single nucleotide polymorphism association with diseases.

Disorders	Adjusted P-value	Combined Score	Genes
Parkinson Disease	0.013661611	13.48649405	KANSL1; C10ORF3
Platelet Aggregation	0.033693075	8.854691276	IL1R2
Brain dysfunction	0.033693075	8.488417033	KANSL1
Behcet Syndrome	0.043550841	6.487656658	IL1R2
Coronary Artery Disease	0.116327628	4.393806844	FRRS1L

are the first rationally designed candidate factors for PD evaluation using blood cell transcripts, but clinical investigations in PD patients are needed to evaluate their utility; nevertheless, their identification employed informatics based analytical logic and large gene expression datasets. What is novel is our approach to find disease markers that are can be employed in a tissue (blood) that is easily accessible in the clinic and does not require specialised methods of evaluation beyond gene expression analysis methods. The nature of these biomarkers and the pathways they participate in may reveal new aspects of PD development and progression, particularly since these biomarkers are evident in cells outside the central nervous system, and so reflect responses to systemic (i.e., note brain-restricted) factors or processes affected by PD. Whether or not these genes have practical utility for diagnostic assessment (which would need proper validation) or may identify pathological processes of interest it seems unlikely that they could form the basis for therapeutic developments since they do not have target tissue specificity.

5. Conclusions

In conclusion, our approach in identifying potential blood markers for PD has identified gene expression markers with potential for diagnostic utility, using blood transcript levels to monitor PD progression. This approach could be thus used to develop blood cell transcript-based tests for other neurodegenerative diseases that cannot yet be evaluated without detailed brain scans or surgically invasive interventions.

Data availability

RNAseq and gene expression microarray datasets related to the blood and brain cells are available at the Gene Expression Omnibus of the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/geo/>) with accession numbers GSE68719 and GSE22491. GWAS data is freely accessible in the NHGRI-EBI Catalog (<https://www.ebi.ac.uk/gwas/>) and eQTL data (both blood and brain) is available in the GTExPortal (<https://gtexportal.org/home/>). Pathways, ontology and PPI datasets are available respectively in the DAVID bioinformatics resources (<https://david-d.ncicrf.gov/>), KEGG pathways and STRING databases [string - db.org](http://string-db.org). Moreover, our incorporated gold benchmark verified datasets is freely accessible from the dbGaP (www.ncbi.nlm.nih.gov/gap).

References

- [1] C. Soto, S. Pritzkow, Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases, *Nat. Neurosci.* 1 (2018).
- [2] M. R. Rahman, T. Islam, T. Zaman, M. Shahjaman, M. R. Karim, F. Huq, J. M. Quinn, R. D. Holsinger, E. Gov, M. A. Moni, Identification of molecular signatures and pathways to identify novel therapeutic targets in alzheimer's disease: insights from a systems biomedicine perspective, *Genomics*. <https://doi.org/10.1016/j.ygeno.2019.07.018>.
- [3] M.G. Erkkonen, M.-O. Kim, M.D. Geschwind, Clinical neurology and epidemiology of the major neurodegenerative diseases, *Cold Spring Harbor Perspect. Biol.* 10 (4) (2018) a033118.
- [4] T. Islam, M.R. Rahman, M.R. Karim, F. Huq, J.M. Quinn, M.A. Moni, Detection of multiple sclerosis using blood and brain cells transcript profiles: insights from comprehensive bioinformatics approach, *Inform. Med. Unlocked* (2019) 100201.
- [5] A. Cieslak, E.E. Smith, J. Lysack, Z. Ismail, Case series of mild behavioral impairment: toward an understanding of the early stages of neurodegenerative diseases affecting behavior and cognition, *Int. Psychogeriatr.* 30 (2) (2018) 273–280.
- [6] M. Rahman, T. Islam, M. Shahjaman, T. Zaman, H.M. Faruquee, M.A.H.M. Jamal, F. Huq, J.M. Quinn, M.A. Moni, et al., Discovering biomarkers and pathways shared by alzheimers disease and ischemic stroke to identify novel therapeutic targets, *Medicina* 55 (5) (2019) 191.
- [7] T. Nonaka, M. Masuda-Suzukake, M. Hasegawa, Molecular mechanisms of the co-deposition of multiple pathological proteins in neurodegenerative diseases, *Neuropathology* 38 (1) (2018) 64–71.
- [8] G.S. Griesbach, B.E. Masel, R.E. Helvie, M.J. Ashley, The impact of traumatic brain injury on later life: effects on normal aging and neurodegenerative diseases, *J. Neurotrauma* 35 (1) (2018) 17–24.
- [9] H.K. Rana, M.R. Akhtar, M.B. Ahmed, P. Liò, J.M. Quinn, F. Huq, M.A. Moni, Genetic effects of welding fumes on the progression of neurodegenerative diseases, *Neurotoxicology* 71 (2019) 93–101, <https://doi.org/10.1016/j.neuro.2018.12.002>.
- [10] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorf, et al., The nhgri gwas catalog, a curated resource of snp-trait associations, *Nucleic Acids Res.* 42 (D1) (2013) D1001–D1006.
- [11] P.M. Visscher, N.R. Wray, Q. Zhang, P. Sklar, M.I. McCarthy, M.A. Brown, J. Yang, 10 years of gwas discovery: biology, function, and translation, *Am. J. Hum. Genet.* 101 (1) (2017) 5–22.
- [12] L.R. Lloyd-Jones, A. Holloway, A. McRae, J. Yang, K. Small, J. Zhao, B. Zeng, A. Bakshi, A. Metspalu, M. Dermizakis, et al., The genetic architecture of gene expression in peripheral blood, *Am. J. Hum. Genet.* 100 (2) (2017) 228–237.
- [13] F.A. Wright, P.F. Sullivan, A.I. Brooks, F. Zou, W. Sun, K. Xia, V. Madar, R. Jansen, W. Chung, Y.-H. Zhou, et al., Heritability and genomics of gene expression in peripheral blood, *Nat. Genet.* 46 (5) (2014) 430.
- [14] G. Consortium, et al., Genetic effects on gene expression across human tissues, *Nature* 550 (7675) (2017) 204.
- [15] A. Rzhetsky, D. Wajngurt, N. Park, T. Zheng, Probing genetic overlap among complex human phenotypes, *Proc. Natl. Acad. Sci.* 104 (28) (2007) 11694–11699.
- [16] D.-S. Lee, J. Park, K. Kay, N.A. Christakis, Z. Olvtvai, A.-L. Barabási, The implications of humanmetabolic network topology for disease comorbidity, *Proc. Proceedings of the National Academy of Sciences* 105 (29) (n.d.) 9880–9885.
- [17] M. Moni, P. Liò, Infectome, diseasome and comorbidities of zika infection, *Int. J. Infect. Dis.* 53 (2016) 14.
- [18] K.-I. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, A.-L. Barabási, The human disease network, *Proc. Natl. Acad. Sci.* 104 (21) (2007) 8685–8690.
- [19] M.A. Moni, P. Liò, Genetic profiling and comorbidities of zika infection, *J. Infect. Dis.* 216 (6) (2017) 703–712.
- [20] I. Feldman, A. Rzhetsky, D. Vitkup, Network properties of genes harboring inherited disease mutations, *Proc. Natl. Acad. Sci.* 105 (11) (2008) 4323–4328.
- [21] H. Xu, M.A. Moni, P. Liò, Network regularised cox regression and multiplex network models to predict disease comorbidities and survival of cancer, *Comput. Biol. Chem.* 59 (2015) 15–31.
- [22] K. Lage, E.O. Karlberg, Z.M. Størling, P.I. Olason, A.G. Pedersen, O. Rigina, A.M. Hinsby, Z. Tümer, F. Pociot, N. Tommerup, et al., A human phenome-interactome network of protein complexes implicated in genetic disorders, *Nat. Biotechnol.* 25 (3) (2007) 309.
- [23] M.A. Hossain, T.A. Asa, J.M. Quinn, M.M. Rahman, F. Huq, M.A. Moni, Network-based Genetic Profiling, and Therapeutic Target Identification of Thyroid Cancer, *bioRxiv*, 2018480632.
- [24] S. Suthram, J.T. Dudley, A.P. Chiang, R. Chen, T.J. Hastie, A.J. Butte, Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets, *PLoS Comput. Biol.* 6 (2) (2010) e1000662.
- [25] M.A. Hossain, T.A. Asa, F. Huq, J.M. Quinn, M.A. Moni, A network-based approach to identify molecular signatures and comorbidities of thyroid cancer, *Proceedings of International Joint Conference on Computational Intelligence*, Springer, 2020, pp. 235–246.
- [26] V. Hachinski, C. Iadecola, R.C. Petersen, M.M. Breteler, D.L. Nyenhuis, S.E. Black, W.J. Powers, C. DeCarli, J.G. Merino, R.N. Kalaria, et al., National institute of neurological disorders and stroke—canadian stroke network vascular cognitive impairment harmonization standards, *Stroke* 37 (9) (2006) 2220–2241.
- [27] Parkinson's foundation, accessed: 2018-09-22, <http://www.parkinson.org/understanding-parkinsons/causes-and-statistics>, (2018).
- [28] D.W. Dickson, Parkinsons Disease and Parkinsonism: *Neuropathology*, Cold Spring Harbor Perspectives in Medicine, (2012) a009258.
- [29] M.H. Rahman, S. Peng, C. Chen, P. Liò, M.A. Moni, Genetic Effect of Type 2 Diabetes to the Progression of Neurological Diseases, *bioRxiv*, 2018480400.
- [30] S. Sveinbjornsdottir, The clinical symptoms of Parkinson's disease, *J. Neurochem.* 139 (2016) 318–324.
- [31] N. Sakib, U.N. Chowdhury, M.B. Islam, J.M. Quinn, M.A. Moni, A System Biology Approach to Identify the Genetic Markers to the Progression of Parkinson's Disease for Aging, Lifestyle and Type 2 Diabetes, *bioRxiv*, 2018482760.
- [32] A. Ashoori, D.M. Eagleman, J. Jankovic, Effects of auditory rhythm and music on gait disturbances in Parkinsons disease, *Front. Neurol.* 6 (2015) 234.
- [33] U.N. Chowdhury, M.B. Islam, S. Ahmad, F. Huq, J.M. Quinn, M.A. Moni, Network-based Identification of Genetic Factors in Ageing, Lifestyle and Type 2 Diabetes that Influence in the Progression of Alzheimer's Disease, *bioRxiv*, 2018482844.
- [34] J.E. Ahlskog, Does vigorous exercise have a neuroprotective effect in Parkinson disease? *Neurology* 77 (3) (2011) 288–294.

- [35] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, Kegg for integration and interpretation of large-scale molecular data sets, *Nucleic Acids Res.* 40 (D1) (2011) D109–D114.
- [36] J.S. Amberger, A. Hamosh, Searching online mendelian inheritance in man (omim): a knowledgebase of human genes and genetic phenotypes, *Curr. Protoc. Bioinform.* 58 (1) (2017) 1–2.
- [37] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for rna-seq data with deseq2, *Genome Biol.* 15 (12) (2014) 550.
- [38] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, Limma powers differential expression analyses for rna-sequencing and microarray studies, *Nucleic Acids Res.* 43 (7) (2015) e47–e47.
- [39] M.A. Moni, P. Liò, Network-based analysis of comorbidities risk during an infection: sars and hiv case studies, *BMC Bioinf.* 15 (1) (2014) 333.
- [40] M.R. Rahman, T. Islam, B. Turanli, T. Zaman, H.M. Faruquee, M.M. Rahman, M.N.H. Mollah, R.K. Nanda, K.Y. Arga, E. Gov, et al., Network-based approach to identify molecular signatures and therapeutic agents in alzheimers disease, *Comput. Biol. Chem.* 78 (2019) 431–439.
- [41] M.A. Moni, H. Xu, P. Liò, Cytocom: a cytoscape app to visualize, query and analyse disease comorbidity networks, *Bioinformatics* 31 (6) (2014) 969–971.
- [42] M.A. Moni, P. Liò, comoR: a software for disease comorbidity risk assessment, *J. Clin. Bioinform.* 4 (1) (2014) 8.
- [43] M.A. Moni, P. Liò, How to build personalized multi-omics comorbidity profiles, *Front. Cell. Dev. Biol.* 3 (2015) 28.
- [44] H.K. Rana, M.R. Akhtar, M.B. Islam, M.B. Ahmed, P. Liò, J.M. Quinn, F. Huq, M.A. Moni, Genetic effects of welding fumes on the development of respiratory system diseases, *Comput. Biol. Med.* 108 (2019) 142–149.
- [45] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, et al., String v10: protein–protein interaction networks, integrated over the tree of life, *Nucleic Acids Res.* 43 (D1) (2014) D447–D452.
- [46] M.R. Rahman, T. Islam, F. Huq, J.M. Quinn, M.A. Moni, Identification of molecular signatures and pathways common to blood cells and brain tissue of amyotrophic lateral sclerosis patients, *Inform. Med. Unlocked* 16 (2019) 100193.
- [47] M. McKenzie, A.K. Henders, A. Caracella, N.R. Wray, J.E. Powell, Overlap of expression quantitative trait loci (eqtl) in human brain and blood, *BMC Med. Genomics* 7 (1) (2014) 31.
- [48] M.A. Moni, H.K. Rana, M.B. Islam, M.B. Ahmed, M.A.M. Hasan, F. Huq, J. Quinn, Early Detection of Neurological Dysfunction Using Blood Cell Transcript Profiles, *bioRxiv*, 2018483016.
- [49] Y. Li, L. Jiang, X. Chen, Y. He, G. Li, Y. Zhang, Discovery of potential atp-sensitive potassium channel openers with potential hypotensive activity from Chinese herbs based on molecular simulation, *Zhongguo Zhong yao za zhi = Zhongguo zhongyao zazhi = China journal of Chinese materia medica* 41 (2) (2016) 264–271.
- [50] M.R. Rahman, T. Islam, M. Shahjaman, J.M. Quinn, R.D. Holsinger, M.A. Moni, Identification of Common Molecular Biomarker Signatures in Blood and Brain of Alzheimers Disease, *BioRxiv*, 2019482828.
- [51] J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, et al., The genotype-tissue expression (gtex) project, *Nat. Genet.* 45 (6) (2013) 580.
- [52] M. Fagny, J.N. Paulson, M.L. Kuijjer, A.R. Sonawane, C.-Y. Chen, C.M. Lopes-Ramos, K. Glass, J. Quackenbush, J. Platig, Exploring regulation in tissues with eqtl networks, *Proc. Natl. Acad. Sci.* 114 (37) (2017) E7841–E7850.
- [53] M.R. Rahman, T. Islam, T. Zaman, M. Shahjaman, M.R. Karim, M.A. Moni, Blood-based Molecular Biomarker Signatures in Alzheimers Disease: Insights from Systems Biomedicine Analyses, *BioRxiv*, 2018481879.