Heliyon 7 (2021) e08518

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Immune gene network of neurological diseases: Multiple sclerosis (MS), Alzheimer's disease (AD), Parkinson's disease (PD) and Huntington's disease (HD)

Shradha Mukherjee^{*,1}

Calcutta, West Bengal, India

ARTICLE INFO

Neurodegenerative diseases

Keywords:

Multiple sclerosis

Brain diseases

Bioinformatics

Gene networks

Gene expression

Immune system

WGCNA

Biomarker RNA-sea

Microglia

Monocyte

Macrophage

ABSTRACT

Neurological diseases, such as MS, AD, PD and HD, are a major health concern of the elderly population, but still therapeutic options are limited. Recent advances in genomic sequencing and bioinformatics, present an opportunity to understand mechanisms of these diseases for identification of therapeutic targets. Several studies have shown association of immune dysfunction with immune system mediated neurological disease, MS, as well as neurodegenerative diseases (AD, PD and HD). However, similarities and differences in role of the immune system, immune pathways and immune cell types in these diseases remains unknown. In this study, immune cell type signature genes in gene networks associated with neurological diseases, MS, AD, PD and HD was investigated using meta-analysis and bioinformatics methods. Application of Weighted Gene Co-expression Network Analysis (WGCNA) on publicly available gene expression datasets (microarray and RNA-seq) revealed a ModArray_04 module (microarray) or ModRNAseq_06 module (RNA-seq), significantly associated with MS, AD, PD and HD. Hypergeometric enrichment test revealed significant enrichment of immune cell type genes in these neurological disease modules. This study demonstrates that immune system mediated neurological disease, MS and neurodegenerative diseases (AD, PD and HD), share a common gene network characterized by immune cell type signature genes (microglia, monocytes and macrophages) and are probable targets for therapeutic intervention. In summary, this work shows a connection between MS, a disease where the role of the immune system and inflammation is established, and neurodegenerative diseases (AD, PD and HD) where the role of inflammation is still a hypothesis.

1. Introduction

It is estimated that number of adults above 65 years of age will outnumber children in U.S. by 2035. This presents a significant challenge to healthcare because age increases susceptibility to life changing neurodegenerative diseases and cancers [1]. Alzheimer's disease (AD) is characterized by appearance of extracellular amyloid-beta plaques, intracellular neurofibrillary tangles and neuronal cell death particularly in cortex, and hippocampus [2]. AD is most common form of dementia occurring in elderly people and overall prevalence of AD is expected to double in next 20 years [3]. Aging is risk factor for Parkinson's disease (PD) and its hallmarks are neurodegeneration in substantia nigra and sever motor dysfunction [4]. PD effects 1% of U.S. population above age 60 and increases to 5% in individuals above age 85 [4]. Huntington's disease (HD) is an inherited neurodegenerative disease caused by unstable CAG trinucleotide repeat expansion in huntingtin (HTT) gene and manifests as cognitive, psychiatric and/or movement dysfunction [5, 6]. Adult onset HD is most common form of this disease and can occur between age 2 to 85 [7]. Therefore, aging is a major risk factor for neurodegenerative diseases, AD, PD and HD.

Several studies have sought to identify mechanistic links between aging and neurodegenerative diseases. The goal of this introduction is to provide readers with an overview of all known mechanisms of AD, PD, HD and MS. Mitochondrial dysfunction, immune system and inflammation are emerging as crucial mechanistic links between aging and neurodegeneration. Mitochondria are cellular organelles that produce cellular energy in form of ATP by oxidative phosphorylation

https://doi.org/10.1016/j.heliyon.2021.e08518

Received 9 October 2021; Received in revised form 18 November 2021; Accepted 29 November 2021





CelPress

* Corresponding author.

E-mail address: smukher2@gmail.com.

¹ This work is self-owned and not affiliated with any organization/university.

^{2405-8440/© 2021} The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/bync-nd/4.0/).

of fats, and carbohydrates in mitochondria [8]. Reduced size and number of mitochondria, altered mitochondrial oxidative phosphorylation, and altered mitochondria morphology occur in AD, PD and HD [9, 10, 11]. 'Neuroinflammation hypothesis' suggests dysfunction of immune system response and inflammation in central nervous system as a driver of neurodegenerative diseases [12]. Microglia, the resident immune cells of brain, are widely studied in context of neurodegeneration and aging. Several studies have shown that microglia signature genes are altered in three neurodegenerative diseases, AD, PD and HD, in context of aging [13, 14, 15, 16, 17, 18]. Due to increased leakiness of blood brain barrier and systemic alterations in immune system with aging, immune cells from peripheral blood enter the brain and influence brain inflammation [19, 20]. However, the role played by different immune cell types in neurodegenerative diseases needs further investigation. Especially, it will be interesting to tease out these regulatory immune mechanisms in context of neurodegeneration independent of aging.

Multiple Sclerosis (MS) is most common chronic inflammatory brain disease mediated by immune system. MS is characterized by demyelination, and axonal degeneration [21]. MS typically occurs in adults between 20 to 40 years of age and is diagnosed based on demyelination lesions [22]. Unlike neurodegenerative diseases, such as AD, PD and HD, old age does not necessarily increase risk of MS. Though no driver gene(s) is known to cause MS, studies have shown that genetic predisposition and environmental factors increase risk of MS [23]. Consistent with immune origin of MS, early life infection by Epstein-Barr virus (EBV) or human herpesvirus 4 and genetic variants of immune genes increase risk of MS [24, 25]. Chronic activation of immune pathway in neurodegenerative diseases (AD, PD and HD) is well known. However, these studies were done on one disease at a time, which makes it challenging to compare the role of immune system across all neurodegenerative diseases (AD, PD and HD) [26, 27, 28, 29]. Two studies Durrenberger et al and Mukherjee et al, compared multiple neurodegenerative diseases and found a neurodegenerative disease associated immune pathway, and immune cell type (microglia), respectively [13, 30]. However, the role played by other immune cells, such as innate immune cell types, monocytes and macrophages, in neurodegenerative diseases (AD, PD and HD) and how it compares to classic immune system mediated disease, MS, remains unknown.

In this work, a side-by-side comparison was made between neurodegenerative diseases (AD, PD and HD) and immune system disorder, MS. Bioinformatics analysis revealed that three immune cell types, microglia, monocytes and macrophages, underlie neurodegenerative diseases (AD, PD and HD) and immune dysfunction mediated neurological disease MS, independent of age.

2. Methods

2.1. Code availability

Computational pipeline used in this work is available on github https: //github.com/smukher2/GithubSubmittedADPDHDMSimmun eSept2021.

2.2. Source of RNA-seq and microarray samples

Publicly available datasets on NIH GEO https://www.ncbi.nlm.nih .gov/geo/ were used in this study. Human microarray datasets were obtained from cortex (control and AD), substantia nigra (control and PD) and whole brain (control and MS). Human RNA-seq datasets were obtained from prefrontal cortex (control, AD, HD, PD), optic chiasm (control and MS) and hippocampus (control and AD). Mouse RNA-seq datasets were used to identify brain cell type genes (astrocytes, neurons, oligodendrocytes, microglia) and blood immune cell type genes (monocytes and macrophages). The series numbers for these datasets are provided at the end of this manuscripts under acknowledgements.

2.3. RNA-seq raw reads processing

RNA-seq analysis pipeline was adapted and developed from previously published standard methodologies [13, 31]. Briefly, RNA-seq raw reads were aligned to GRCh38 and GRCm38 genomes with TopHat using gene annotation information from Homo_sapiens.GRCh38.92.gtf and Mus_musculus.GRCm38.92.gtf files, for human datasets (AD, PD, HD and MS), and mouse datasets (brain cell types and immune cell types), respectively, on cyverse high-performance computational cluster [32]. Mapped output bam files were used as input for HTSeq-Count to count reads overlapping genes for estimation of gene abundance [33]. All further analyses, including graphical visualizations were done in R environment with R packages and custom R functions. Raw gene counts were converted to TPM values (filtered to keep values > 1 in at least 2 samples) to normalized for variations in library size and mRNA abundance. TPM values were further scaled to log2TPM+1 scale and used as input for rest of the bioinformatics pipeline. The log2TPM+1 values were visualized with volcano plots, barplots and density plots using ggplot2_3.1.1 R package [34]. Built-in R functions, prcomp stats4_3.5.0 and corrplot 0.84, were used to perform PCA analyses and correlation analysis, respectively, to visualize variability of replicate samples before and after TPM normalization [34, 35].

2.4. Microarray data processing

Previously published microarray analysis pipeline was used here [13]. GEOquery R package was used to obtain microarray datasets and GPL annotation files [36]. Expression values from all datasets were quantile normalized using limma R package and converted to log2+1 scale [37]. Probe IDs were converted to gene symbols and in case of duplicates, the first probe ID gene expression value was kept for all datasets. The log2+1 expression values were used as input for estimation of DEGs (Differentially Expressed Genes) and construction of gene co-expression networks (WGCNA network). The log2+1 expression values were visualized with volcano plots, barplots and density plots using ggplot2_3.1.1 R package [34].

2.5. Consensus differential gene expression detection from limma, edgeR and simple comparison of means

Differentially expressed genes (DEGs) between a given cell type and all other cell types, were identified using limma_3.38.3, edgeR_3.24.3 and simple comparison expression means [37, 38, 39]. In limma and edgeR, model design included condition, study or batch, gender, age and tissue. For calculation of DEGs in limma and edgeR, contrasts were made for the two conditions being compared holding all other variables in design model constant. Limma and edgeR methods included correction for false discovery and multiple testing using Benjamini and Hochberf (BH) to correct for the presence of >30K transcripts (genes, pseudogenes and antisense) in the analysis [40]. Though DEG analysis outputs both upregulated and downregulated genes, here only upregulated genes were selected as cell type specific signature genes. Signature or marker genes were defined by significant presence of genes (upregulation) and not significant absence of genes (downregulation) because generally cells are identified by presence of marker genes for techniques such as, Fluorescence-activated cell sorting and immunohistochemistry. BH corrected adjP-values <0.05 and a fold changes >5 were cut-offs used for significantly upregulated genes. Gene expression of DEGs were visualized with volcano plots, barplots and density plots using ggplot2_3.1.1 R package [34]. Consensus DEGs were those genes that were present in DEG lists of at least two of the three methods, limma, edgeR and simple comparison of means. Significance of overlap between DEG lists were calculated using GeneOverlap_1.18.0 R package and number of genes overlapping were visualized as Venn-diagrams using VennDiagram_1.6.20 R package [41, 42]. UserListEnrichment R function was



Figure 1. Effect of SVA + LM adjustment on AD, PD, HD, MS and control brain microarray datasets (GSE33000, GSE20333, GSE26927, GSE20164, GSE20292, GSE108000 and GSE43490). A: Density plot representation of gene expression per study before and after SVA + LM adjustment B: Box and whisker plot representation of gene expression per study before and after SVA + LM adjustment. C: Principal component (PC) projection of datasets before and after SVA + LM adjustment. D: Pearson correlation plot of studies before and after SVA + LM adjustment. For this figure the following R packages were used: prcomp built-in R function in stats4_3.5.0, corrplot_0.84, ggplot2_3.1.1, sva_3.30.1 and limma_3.38.3 [13,34,35,37,45, 46, 47].



Figure 2. Effect of SVA + LM adjustment on AD, PD, HD, MS and control brain RNA-seq datasets (GSE53697, GSE64810, GSE68719, GSE100297 and GSE67333). A: Density plot representation of gene expression per study before and after SVA + LM adjustment B: Box and whisker plot representation of gene expression per study before and after SVA + LM adjustment. C: Principal component (PC) projection of datasets before and after SVA + LM adjustment. D: Pearson correlation plot of studies before and after SVA + LM adjustment. For this figure the following R packages were used: prcomp built-in R function in stats4_3.5.0, corrplot_0.84, ggplot2_3.1.1, sva_3.30.1 and limma_3.38.3 [13,34,35,37,45, 46, 47].

Table 1	1. Number	of genes	in microarray	and RNA-seq	WGCNA modules.
---------	-----------	----------	---------------	-------------	----------------

1ModArray.0121411ModRNAseq.0132402ModArray.0214862ModRNAseq.0218503ModArray.037833ModRNAseq.0316904ModArray.043354ModRNAseq.0416685ModArray.052975ModRNAseq.0516386ModArray.062456ModRNAseq.071778ModArray.071947ModRNAseq.071178ModArray.091729ModRNAseq.0911129ModArray.091729ModRNAseq.09111210ModArray.1113411ModRNAseq.1467512ModArray.1210812ModRNAseq.1261213ModArray.1310713ModRNAseq.1466915ModArray.149914ModRNAseq.1464916ModRNAseq.1466914ModRNAseq.1464915ModArray.159915ModRNAseq.163416ModArray.168017ModRNAseq.163417ModRNAseq.186618ModRNAseq.1929816ModArray.195319ModRNAseq.1929817ModArray.196114ModRNAseq.2129418ModArray.192921ModRNAseq.2329419ModArray.193721ModRNAseq.2329420ModArray.20		WGCNA_microarray _module_names	Number_of_genes		WGCNA_RNA-seq _module_names	Number_of_genes
2 ModAray.02 1486 2 ModRNacq.02 1850 3 ModAray.03 783 3 ModRNacq.04 1669 5 ModAray.05 297 5 ModRNacq.05 1638 6 ModAray.06 245 6 ModRNacq.06 1586 7 ModAray.07 194 7 ModRNacq.06 1177 8 ModAray.08 186 8 ModRNacq.08 1112 9 ModRNacq.07 112 9 ModRNacq.08 112 10 ModAray.09 122 9 ModRNacq.09 112 10 ModAray.10 139 10 ModRNacq.01 35 12 ModAray.11 134 14 ModRNacq.11 81 13 ModAray.13 107 13 ModRNacq.13 64 14 ModAray.13 107 13 ModRNacq.13 64 14 ModAray.13 99 14 ModRNacq.14 64 15 ModAray.15 91 ModRNacq.16 34 16 ModRNacq.16 84 14 14 16 ModRNacq.13 14 14 16 ModRNacq.14 24	1	ModArray_01	2141	1	ModRNAseq_01	3240
3ModAray.037833ModRNaseq.0316904ModAray.043554ModRNaseq.0416685ModAray.052975ModRNaseq.0516386ModAray.062456ModRNaseq.071778ModAray.071947ModRNaseq.0711278ModAray.091729ModRNaseq.09111210ModAray.091729ModRNaseq.1095611ModAray.1113411ModRNaseq.1287512ModAray.1210812ModRNaseq.1371414ModAray.1310713ModRNaseq.1466915ModAray.149914ModRNaseq.1466916ModAray.159914ModRNaseq.1438416ModAray.168116ModRNaseq.1738417ModAray.159914ModRNaseq.1738418ModAray.166618ModRNaseq.1831419ModAray.195319ModRNaseq.1929620ModAray.195220ModRNaseq.2029621ModAray.20227627627622ModAray.20222ModRNaseq.2324424ModAray.2022624424425ModRNaseq.2422292426ModRNaseq.25244242427ModRNase	2	ModArray_02	1486	2	ModRNAseq_02	1850
4ModAray_043354ModRNaseq_0416685ModAray_052975ModRNaseq_0515386ModAray_062456ModRNaseq_071728ModAray_081868ModRNaseq_0811129ModAray_091729ModRNaseq_09111210ModAray_1013910ModRNaseq_1095611ModAray_1113411ModRNaseq_1287512ModAray_1210812ModRNaseq_1371413ModAray_1310713ModRNaseq_136914ModAray_149914ModRNaseq_146915ModAray_15811134434416ModAray_168016ModRNaseq_158117ModAray_178017ModRNaseq_1738418ModAray_186618ModRNaseq_183419ModAray_186618ModRNaseq_1929421ModAray_193721ModRNaseq_2029422ModAray_205220ModRNaseq_212423ModAray_2022ModRNaseq_2424424ModRNaseq_24244242424ModRNaseq_25294242425ModRNaseq_262442442426ModRNaseq_26244242427ModRNaseq_2624424424<	3	ModArray_03	783	3	ModRNAseq_03	1690
5 ModArray.05 297 5 ModArseq.05 1638 6 ModArray.06 245 6 ModRNaceq.06 1586 7 ModArray.07 194 7 ModRNaceq.07 1177 8 ModArray.08 1636 8 ModRNaceq.08 1115 9 ModArray.09 172 9 ModRNaceq.09 112 10 ModArray.10 139 10 ModRNaceq.10 956 11 ModArray.11 134 10 ModRNaceq.11 875 13 ModArray.12 108 12 ModRNaceq.12 812 14 ModArray.13 107 13 ModRNaceq.14 669 15 ModArray.14 99 14 ModRNaceq.14 669 16 ModArray.16 80 16 ModRNaceq.14 84 17 ModArray.16 80 16 ModRNaceq.14 94 18 ModArray.16 80 16 ModRNaceq.14 94 19 ModArray.16 80 16 ModRNaceq.14 94 19 ModArray.16 66 18 ModRNaceq.14 94 19 ModArray.02 52 10	4	ModArray_04	335	4	ModRNAseq_04	1668
6ModAray.062456ModRNAseq.0615867ModAray.071947ModRNAseq.0711778ModAray.091868ModRNAseq.0811159ModAray.091729ModRNAseq.1095611ModAray.1013910ModRNAseq.1187512ModAray.1210812ModRNAseq.1281213ModAray.1310713ModRNAseq.1371414ModAray.149914ModRNAseq.146915ModAray.159914ModRNAseq.1681416ModAray.168016ModRNAseq.1738417ModAray.178017ModRNAseq.1631418ModAray.186618ModRNAseq.1738419ModAray.195319ModRNAseq.132920ModAray.205220ModRNAseq.2127621ModAray.202922ModRNAseq.212923ModAray.002920ModRNAseq.222424ModRNAseq.26242425ModRNAseq.26202426ModRNAseq.26202427ModRNAseq.26242428ModRNAseq.26242429ModRNAseq.26242420ModRNAseq.26242421ModRNAseq.26242423ModRNAseq.26<	5	ModArray_05	297	5	ModRNAseq_05	1638
7ModArray.071947ModRNAseq.0711778ModArray.081868ModRNAseq.0811159ModArray.091729ModRNAseq.0911210ModArray.1013910ModRNAseq.1095611ModArray.1113411ModRNAseq.1281212ModArray.1210812ModRNAseq.1371413ModArray.1310713ModRNAseq.1371414ModArray.149914ModRNAseq.1558115ModArray.159915ModRNAseq.1558116ModArray.168016ModRNAseq.1738417ModArray.178017ModRNAseq.1929218ModArray.186618ModRNAseq.1929420ModArray.095220ModRNAseq.2029421ModArray.09220ModRNAseq.2127622ModArray.00222ModRNAseq.2424423ModArray.00223ModRNAseq.2422424ModRNAseq.24224242425ModRNAseq.24224242426ModRNAseq.24224242427ModRNAseq.24224242428ModRNAseq.2620242429ModRNAseq.2620242429ModRNAseq.26242424<	6	ModArray_06	245	6	ModRNAseq_06	1586
8 ModArray.08 186 8 ModRNAseq.08 1115 9 ModArray.09 172 9 ModRNAseq.09 1112 10 ModArray.10 139 10 ModRNAseq.10 956 11 ModArray.11 134 ModRNAseq.12 812 13 ModArray.12 108 13 ModRNAseq.12 812 14 ModArray.13 107 13 ModRNAseq.14 669 15 ModArray.14 99 14 ModRNAseq.14 669 16 ModRNAseq.14 69 51 51 16 ModRNAseq.14 69 32 34 16 ModRNAseq.17 384 34 17 ModArray.16 80 16 ModRNAseq.17 384 18 ModArray.18 66 18 ModRNAseq.19 34 19 ModArray.19 53 19 ModRNAseq.20 29 21 ModArray.20 52 20 ModRNAseq.21 24 22 ModArray.00 29 24 24 23 ModArray.20 29 24 24 24 ModRNAseq.24 24 24 25 <td>7</td> <td>ModArray_07</td> <td>194</td> <td>7</td> <td>ModRNAseq_07</td> <td>1177</td>	7	ModArray_07	194	7	ModRNAseq_07	1177
9ModArray.091729ModRNaseq.09111210ModArray.1013910ModRNaseq.1095611ModArray.1113411ModRNaseq.1187512ModArray.1210812ModRNaseq.1281213ModArray.1310713ModRNaseq.1466915ModArray.149914ModRNaseq.1558116ModArray.168016ModRNaseq.1642417ModArray.168017ModRNaseq.1738418ModArray.178019ModRNaseq.1831419ModArray.186618ModRNaseq.1929820ModArray.205220ModRNaseq.2127621ModArray.202920ModRNaseq.2122922ModArray.202022ModRNaseq.2322923ModArray.00223ModRNaseq.2422424ModRNaseq.25224242425ModRNaseq.26200242426ModRNaseq.2718636362712ModRNaseq.2620202028ModRNaseq.2718636363629ModRNaseq.2810536363629ModRNaseq.2810536363629ModRNaseq.2810536363629ModRNaseq.28105 <td< td=""><td>8</td><td>ModArray_08</td><td>186</td><td>8</td><td>ModRNAseq_08</td><td>1115</td></td<>	8	ModArray_08	186	8	ModRNAseq_08	1115
10ModArray.1013910ModRNAseq.1095611ModArray.1113411ModRNAseq.1187512ModArray.1210812ModRNAseq.1281213ModArray.1310713ModRNAseq.1371414ModArray.139914ModRNAseq.1466915ModRNAseq.146915ModRNAseq.1558116ModArray.159916ModRNAseq.1642417ModArray.168016ModRNAseq.1738418ModArray.186618ModRNAseq.1831419ModArray.195319ModRNAseq.1929820ModArray.205220ModRNAseq.2029421ModArray.202922ModRNAseq.2124423ModArray.202923ModRNAseq.2422424ModRNAseq.24224242425ModRNAseq.252242426ModRNAseq.262002027ModRNAseq.2718620028ModRNAseq.28200202429ModRNAseq.28200202429ModRNAseq.28200202429ModRNAseq.28200202429ModRNAseq.28200202429ModRNAseq.28200202429ModRNAseq.2820020262	9	ModArray_09	172	9	ModRNAseq_09	1112
11ModArray.1113411ModRNAseq.1187512ModArray.1210812ModRNAseq.1281213ModArray.1310713ModRNAseq.1371414ModArray.149914ModRNAseq.1466915ModArray.159915ModRNAseq.1558116ModArray.168016ModRNAseq.1642417ModArray.168017ModRNAseq.1831418ModArray.186618ModRNAseq.1929820ModArray.195319ModRNAseq.2029421ModArray.202920ModRNAseq.2329422ModArray.202922ModRNAseq.2329423ModArray.00223ModRNAseq.2422424ModRNAseq.2522424424425ModRNAseq.2620424426ModRNAseq.272824427ModRNAseq.2422424428ModRNAseq.2422424429ModRNAseq.2524424429ModRNAseq.2620424429ModRNAseq.2718624429ModRNAseq.2718624429ModRNAseq.2820424429ModRNAseq.2820424429ModRNAseq.2620424429ModRNAseq.2718624429ModRNAseq.28	10	ModArray_10	139	10	ModRNAseq_10	956
12ModArray.1210812ModRNAseq.1281213ModArray.1310713ModRNAseq.1371414ModArray.149914ModRNAseq.1466915ModArray.159915ModRNAseq.1658116ModArray.168016ModRNAseq.1738417ModArray.176618ModRNAseq.1831418ModArray.186618ModRNAseq.1929820ModArray.195319ModRNAseq.2029421ModArray.202920ModRNAseq.2127622ModArray.202922ModRNAseq.2329223ModArray.00232ModRNAseq.242442455ModRNAseq.2522424425ModRNAseq.2622424426ModRNAseq.2624424427ModRNAseq.2624424428ModRNAseq.2624424429ModRNAseq.2624424429ModRNAseq.2624424429ModRNAseq.2624424429ModRNAseq.2624424429ModRNAseq.2624424429ModRNAseq.2624424429ModRNAseq.2624424429ModRNAseq.2624424429ModRNAseq.2624424429ModRNAseq.27186244 <t< td=""><td>11</td><td>ModArray_11</td><td>134</td><td>11</td><td>ModRNAseq_11</td><td>875</td></t<>	11	ModArray_11	134	11	ModRNAseq_11	875
13ModArray.1310713ModRNAseq.1371414ModArray.149914ModRNAseq.1466915ModArray.159915ModRNAseq.1558116ModArray.168016ModRNAseq.1642417ModArray.178017ModRNAseq.1738418ModArray.186618ModRNAseq.1929820ModArray.195319ModRNAseq.2929421ModArray.205220ModRNAseq.2127622ModArray.213721ModRNAseq.2224723ModArray.00223ModRNAseq.2322924SModRNAseq.2422424425ModRNAseq.2522424426ModRNAseq.2620024427ModRNAseq.2620024428ModRNAseq.2718624429ModRNAseq.2620024429ModRNAseq.2620024420ModRNAseq.2620024420ModRNAseq.2620024420ModRNAseq.2620024420ModRNAseq.2620024420ModRNAseq.2620024420ModRNAseq.2620024420ModRNAseq.2620024420ModRNAseq.2620024420ModRNAseq.2620024420ModRNAseq.26 <t< td=""><td>12</td><td>ModArray_12</td><td>108</td><td>12</td><td>ModRNAseq_12</td><td>812</td></t<>	12	ModArray_12	108	12	ModRNAseq_12	812
14ModArray.149914ModRNAseq.1466915ModArray.159915ModRNAseq.1558116ModArray.168016ModRNAseq.1642417ModArray.178017ModRNAseq.1738418ModArray.186618ModRNAseq.1831419ModArray.195319ModRNAseq.2029420ModArray.205220ModRNAseq.2029421ModArray.213721ModRNAseq.2127623ModArray.00223ModRNAseq.2322924ModArray.00223ModRNAseq.2422425ModRNAseq.2424242426ModRNAseq.25242427ModRNAseq.26202028ModRNAseq.27262629ModRNAseq.28202429ModRNAseq.28202420ModRNAseq.26202421ModRNAseq.26202425ModRNAseq.26202426ModRNAseq.271862629ModRNAseq.281052629ModRNAseq.281052629ModRNAseq.0022429ModRNAseq.0022429ModRNAseq.281052420ModRNAseq.281052420ModRNAseq.00224 <t< td=""><td>13</td><td>ModArray_13</td><td>107</td><td>13</td><td>ModRNAseq_13</td><td>714</td></t<>	13	ModArray_13	107	13	ModRNAseq_13	714
15 ModAray,15 99 15 ModRNAseq,15 581 16 ModAray,16 80 16 ModRNAseq,16 424 17 ModAray,17 80 17 ModRNAseq,17 384 18 ModAray,18 66 18 ModRNAseq,18 314 19 ModAray,19 53 19 ModRNAseq,19 298 20 ModAray,20 52 20 ModRNAseq,20 294 21 ModAray,21 37 21 ModRNAseq,21 276 22 ModAray,22 29 24 24 23 ModAray,20 29 24 24 24 ModRNaseq,23 29 29 23 ModAray,00 2 23 ModRNAseq,24 24 24 S ModRNAseq,25 20 20 24 54 ModRNAseq,26 20 20 24 54 ModRNAseq,27 186 24 54 ModRNAseq,28 20 20 24 54 ModRNAseq,28	14	ModArray_14	99	14	ModRNAseq_14	669
16ModArray168016ModRNAseq.1642417ModArray178017ModRNAseq.1738418ModArray186618ModRNAseq.1831419ModArray195319ModRNAseq.1929820ModArray205220ModRNAseq.2029421ModArray213721ModRNAseq.2127622ModArray.002922ModRNAseq.2224723ModArray.00223ModRNAseq.2322924ModRNAseq.24224242425ModRNAseq.2522420026ModRNAseq.26200212627ModRNAseq.27186262628ModRNAseq.28105282829ModRNAseq.281052926	15	ModArray_15	99	15	ModRNAseq_15	581
17 ModArray_17 80 17 ModRNAseq.17 384 18 ModArray_18 66 18 ModRNAseq.18 314 19 ModArray_19 53 19 ModRNAseq.19 298 20 ModArray_20 52 20 ModRNAseq.20 294 21 ModArray_21 37 21 ModRNAseq.21 276 22 ModArray_20 29 22 ModRNAseq.22 247 23 ModArray_00 2 23 ModRNAseq.24 224 24 ModRNAseq.24 224 24 25 ModRNAseq.25 224 24 26 ModRNAseq.26 200 20 27 ModRNAseq.26 20 24 28 ModRNAseq.27 186 26 29 ModRNAseq.28 105 36 29 ModRNAseq.20 20 36	16	ModArray_16	80	16	ModRNAseq_16	424
18 ModArray,18 66 18 ModRNAseq,18 314 19 ModArray,19 53 19 ModRNAseq,19 298 20 ModArray,20 52 20 ModRNAseq,20 294 21 ModArray,21 37 21 ModRNAseq,21 276 22 ModArray,20 29 22 ModRNAseq,22 247 23 ModArray,00 2 33 ModRNAseq,23 229 24 ModRNAseq,24 224 24 25 ModRNAseq,25 224 26 ModRNAseq,26 200 27 ModRNAseq,27 186 28 ModRNAseq,28 105 29 ModRNAseq,200 2	17	ModArray_17	80	17	ModRNAseq_17	384
19 ModArray_19 53 19 ModRNAseq.19 298 20 ModArray_20 52 20 ModRNAseq.20 294 21 ModArray_21 37 21 ModRNAseq.21 276 22 ModArray_00 29 22 ModRNAseq.22 247 23 ModArray_00 2 23 ModRNAseq.23 229 24 ModRNAseq.24 224 224 25 ModRNAseq.25 224 26 ModRNAseq.26 200 27 ModRNAseq.27 186 28 ModRNAseq.28 105 29 ModRNAseq.28 105	18	ModArray_18	66	18	ModRNAseq_18	314
20 ModArray_20 52 20 ModRNAseq_20 294 21 ModArray_21 37 21 ModRNAseq_21 276 22 ModArray_22 29 22 ModRNAseq_22 247 23 ModArray_00 2 23 ModRNAseq_23 229 24 ModRNAseq_24 224 24 25 ModRNAseq_25 224 26 ModRNAseq_26 200 27 ModRNAseq_27 186 28 ModRNAseq_28 105 29 ModRNAseq_00 2	19	ModArray_19	53	19	ModRNAseq_19	298
21 ModArray_21 37 21 ModRNaseq.21 276 22 ModArray_22 29 24 247 23 ModArray_00 2 37 ModRNaseq.23 229 24 ModRNaseq.24 224 24 24 25 ModRNaseq.25 224 24 26 ModRNaseq.26 200 200 27 ModRNaseq.27 186 186 28 ModRNaseq.28 105 29 29 ModRNaseq.00 2 2	20	ModArray_20	52	20	ModRNAseq_20	294
22 ModArray_22 29 24 23 ModArray_00 2 33 ModRNAseq_23 229 24 ModRNAseq_24 224 24 25 ModRNAseq_25 224 26 ModRNAseq_26 200 27 ModRNAseq_27 186 28 ModRNAseq_28 105 29 ModRNAseq_00 2	21	ModArray_21	37	21	ModRNAseq_21	276
23 ModArray_00 2 23 ModRNAseq_23 229 24 ModRNAseq_24 224 25 ModRNAseq_25 224 26 ModRNAseq_26 200 27 ModRNAseq_27 186 28 ModRNAseq_28 105 29 ModRNAseq_00 2	22	ModArray_22	29	22	ModRNAseq_22	247
24 ModRNAseq_24 224 25 ModRNAseq_25 224 26 ModRNAseq_26 200 27 ModRNAseq_27 186 28 ModRNAseq_28 105 29 ModRNAseq_00 2	23	ModArray_00	2	23	ModRNAseq_23	229
25 ModRNAseq_25 224 26 ModRNAseq_26 200 27 ModRNAseq_27 186 28 ModRNAseq_28 105 29 ModRNAseq_00 2				24	ModRNAseq_24	224
26 ModRNAseq_26 200 27 ModRNAseq_27 186 28 ModRNAseq_28 105 29 ModRNAseq_00 2				25	ModRNAseq_25	224
27 ModRNAseq_27 186 28 ModRNAseq_28 105 29 ModRNAseq_00 2				26	ModRNAseq_26	200
28 ModRNAseq_28 105 29 ModRNAseq_00 2				27	ModRNAseq_27	186
29 ModRNAseq_00 2				28	ModRNAseq_28	105
				29	ModRNAseq_00	2

used to uncover any significant overlap of cell type genes between cell types [29].

2.6. Consensus gene co-expression network detection from RNA-seq and microarray datasets with WGCNA

To build gene co-expression networks or modules of highly coexpressed genes significantly associated with traits of interest WGCNA_1.66 R package was used [43, 44]. To model and remove gene expression effects of covariants, such as study or batch, gender, age, tissue and surrogate variables (hidden variables), and retain only disease condition effects, surrogate variable adjustment and linear model (SVA + LM) methods were applied with sva_3.30.1 and limma_3.38.3 R packages [13, 37, 45, 46, 47]. SVA + LM adjustment and WGCNA analysis were done in parallel on both RNA-seq and microarray datasets. SVA + LM adjusted values were converted to log2+1 normalized gene expression and used as input for WGCNA to construct gene networks or modules based on gene co-expression. In WGCNA, gene expression correlations were determined to calculate a topology overlap matrix (TOM) and hierarchical clustering method was applied to build weighted gene co-expression networks or modules with default parameters and minimum module size 100 [43,44]. WGCNA moduleEigengenes R function was used to represent each module with its own eigengene [43, 44]. While Pearson correlation is good for linear relations between continuous variables, Spearman correlation also works for monotonic relations and ordinal (categorical) variables. As biological interactions are complex, often monotonic and the disease state trait is categorical, here Spearman correlation was used to determine module disease association. To determine module trait association, Spearman correlation was calculated for pairs of module eigengenes and traits. Module eigengenes are computed by the WGCNA package and represent the gene expression pattern for all genes in a given module. Disease covariant or trait was split into pairs of AD-control, PD-control, HD-control and MS-control, to identify modules associated with a given disease independent of other diseases. Consensus modules between RNA-seq and microarray datasets were identified using module preservation and useListEnrichment R functions from WGCNA [29, 43, 44]. Preservation analysis was used to determine if connectivity of genes in microarray modules held true in RNA-seq and significance was reported as a Z-summary score. Z-summary scores higher than 5 are recommended to be significant as per creators of WGCNA package. Only disease associated modules that were preserved between the two platforms (microarray and RNA-seq) (Figure 3E) and had significantly overlapping genes (Figure 3F) were selected. We used this method of selection as irrespective of platform the modules showed disease association. The gene overlap also helps identify the name of the equivalent module in RNA-seq platform given that preservation analysis only takes the input of the reference microarray network module names.

2.7. Characterization of gene co-expression disease associated modules

UserListEnrichment R function was used to determine enrichment of cell type signature genes in WGCNA modules [29]. EnrichR_1.0 a R package was used to determine biological process gene ontology (GO) characterization of the modules [48]. Cell type signature genes underlying only significantly preserved disease associated modules, as determined by preservation analysis in the previous step, were reported in Table 2.





(caption on next page)

34 6.35E-28

43 1.23E-35

42 2.47E-41

15 3.84E-05

ModRNAseg 15

ModRNAseq_03

ModRNAseq_07

ModRNAseq_03

AodArray 16

ModArray 19

AodArray_20

/lodArray_21

Figure 3. Identification of AD, PD, HD and MS associated WGCNA modules in microarray and RNA-seq datasets. A: Bar-plot showing number of genes in AD, PD, HD and MS associated microarray WGCNA modules. B: Spearman correlation coefficients and p-values for significant AD, PD, HD, and/or MS associated microarray WGCNA modules. C: Bar-plot showing number of genes in AD, PD, HD and MS associated RNA-seq WGCNA modules. D: Spearman correlation coefficients and p-values for significant AD, PD, HD and MS associated RNA-seq WGCNA modules. D: Spearman correlation coefficients and p-values for significant AD, PD, HD and MS associated RNA-seq WGCNA modules. E: Preservation analysis between all microarray and RNA-seq WGCNA modules using microarray WGCNA module labels. F: Table showing microarray WGCNA modules associated with AD, PD, HD and MS that significantly overlapped with RNA-seq WGCNA modules. For this figure the following R packages were used: WGCNA_1.66 and ggplot2_3.1.1 [34,43,44].

3. Results

3.1. SVA + LM approach reduces batch effects across AD, PD, HD and MS microarray and RNA-seq studies

To identify neurodegenerative disease specific transcriptome, publicly available control and neurological disease (AD, PD, HD and MS) brain gene expression datasets (microarray and RNA-seq) were normalized in preparation for meta-analysis. SVA + LM model's ability to effectively reduce batch effects or cross study variability in microarray and RNA-seq, was visualized by comparing pre-SVA + LM normalized (before) and SVA + LM normalized (after) expression data plots. Density plot comparison of expression data colored by study, shows a greater overlap, indicating lower variability between studies after SVA + LM normalization (Figure 1A for microarray, Figure 2A for RNA-seq). In Box-Whiskers plot, slightly skewed mean expression in GSE33000 and GSE108000 microarray studies, and GSE67333 RNA-seq study, were fixed by SVA + LM normalization (Figure 1B for microarray, Figure 2B for RNA-seq). Wide dispersion of samples from same study (same color) in PCA plot was reduced by SVA + LM normalization (Figure 1C for microarray, Figure 2C for RNA-seq). Correlation plots used to evaluate correlation of gene expression values between studies, showed increased positive correlation after SVA + LM normalization (Figure 1D for microarray, Figure 2D for RNA-seq). Thus, SVA + LM normalization reduced variability unrelated to disease in AD, PD, HD, MS and control gene expression datasets (microarray and RNA-seq), making the data suitable for metaanalysis.

3.2. WGCNA network analysis reveals a common neurodegenerative disease (AD, PD and HD) and MS specific gene co-expression module

Gene co-expression networks were constructed with WGCNA to uncover molecular mechanisms in neurodegenerative diseases (AD, PD and HD) and MS. WGCNA network analysis of microarray data revealed 23 modules, where ModArray_01 was largest module of size 2141 genes and ModArray_22 was smallest module of size 29 genes (Table1). WGCNA network analysis of RNA-seq data revealed 29 modules, where ModRNAseq_01 was largest module of size 3240 genes and ModRNAseq_28 was smallest module of size 105 genes (Table1). Spearman correlations between module eigengenes and disease trait (AD, PD, HD and MS) in microarray dataset, showed 22 WGCNA modules significantly associated with disease trait (Figure 3A,B). Spearman correlation between module eigengenes and disease trait (AD, PD, HD and MS) in RNA-seq dataset, showed 27 WGCNA modules significantly associated with disease trait (Figure 3C,D). Few WGCNA modules were distinctly associated with only one of four diseases, such as PD only module ModArray_17 (microarray), MS only modules ModRNAseq_04/13/16/23/25/28 (RNA-seq), HD only module ModRNAseq_19 (RNA-seq), AD only module ModRNAseq_20 (RNA-seq) (Figure 3B,D). Widespread 100% preservation of microarray and RNA-seq WGCNA modules was observed (Figure 3E). Hypergeometric enrichment test showed significant overlap of disease specific (AD, PD, HD and MS) WGCNA modules from microarray and RNA-seq datasets (Figure 3F). Thus, microarray and RNA-seq WGCNA gene network analyses complemented each other in identifying AD, PD, HD and MS disease specific modules, that were largely preserved.

3.3. Differential gene expression analysis reveals brain cell type and immune cell type specific marker genes

To identify brain cell type (neuron, astrocyte, oligodendrocyte and endothelial) and immune cell type (microglia, monocyte and macrophage) specific signature genes, differential gene expression metaanalysis was performed with three DEG analysis methods (edgeR, limma and simple comparison of means). Comparison of astrocytes with other cell types revealed 1509 DEGs (fold change >1.25, p-value <0.05) in at least two of the three DEG methods (Figure 4A). Endothelial cells compared to other cell types showed 477 DEGs (fold change >1.25, pvalue <0.05) in at least two of the three DEG methods (Figure 4B). Each immune cell type, when compared with other cell types showed 592 DEGs for macrophages, 344 DEGs for microglia and 863 DEGs for monocytes, with fold change >1.25 and p-value <0.05, in at least two of the three DEG methods (Figure 4C, D, 4E). Oligodendrocytes, when compared with other cell types showed 124 DEGs for mature oligodendrocytes, 128 DEGs for precursor oligodendrocytes and 157 DEGs for new oligodendrocytes, with fold change >1.25 and p-value <0.05, in at least two of the three DEG methods (Figure 4F-H). Comparison of neurons with other cell types revealed 889 DEGs (fold change >1.25, p-value <0.05) in at least two of the three DEG methods (Figure 4I). UserListEnrichment R function showed that only a) new oligodendrocyte and precursor oligodendrocyte b) myelinated oligodendrocytes and precursor oligodendrocyte c) macrophages and monocytes, shared significant number of common signature genes.

3.4. Immune cell type (microglia, macrophage and monocyte) marker genes were enriched in AD, PD, HD and MS associated WGCNA module

Brain cell type (neuron, astrocyte, oligodendrocyte and endothelial) and immune cell type (microglia, monocyte and macrophage) specific marker genes identified by differential gene expression analysis were used to estimate relative enrichment of these cells in AD, PD, HD and MS WGCNA modules. Hypergeometric enrichment analysis revealed cell type enrichment in AD, PD, HD and MS associated WGCNA modules (Figure 5A,B, Table2). ModArray_03 microarray WGCNA module and ModRNAseq_14 RNA-seq_WGCNA module, were most significantly enriched with neuron cell type markers (Table2). Interestingly, Mod-Array_04 microarray WGCNA module and ModRNAseq_06 RNA-seq WGCNA module, were the only modules that were significantly enriched with all three-immune cell type (macrophage, microglia and monocyte) markers (Table2, Figure 5A,B). ModArray_06 microarray WGCNA module and ModRNAseq 08 RNA-seq WGCNA module, were significantly enriched with myelinated oligodendrocyte cell type markers (Table2). ModArray_10 and ModArray_20 microarray WGCNA modules, and ModRNAseq_07 WGCNA module, were significantly enriched with astrocyte cell type markers (Table2). ModArray_14 microarray WGCNA module and ModRNAseq_18 RNA-seq WGCNA module, were significantly enriched with endothelial cell type markers (Table2). Though microarray WGCNA module ModArray_15 was enriched with neuron cell type signature genes, its equivalent ModRNAseq_02 RNA-seq WGCNA module was not significantly enriched with any cell type genes (Table2). Thus, each brain cell type occupies separate WGCNA modules, while all three immune cell types (macrophage, microglia and monocyte) occupy the same WGCNA module. This suggests that mechanistically each brain cell type utilizes a separate gene network or module, while all immune cell types utilize the same gene network or module.

358 DEG

DEG_I

N.S.: Not Significant; ---

DEG

Ignored

DEG



Figure 4. A–I: Venn Diagram gene overlap and heat-map of overlap significance for differential gene expression calculation methods (limma, edgeR and simple comparison of means). Upregulated DEGs for contrasts Astrocytes by others (A), Endothelial cells by others (B), Macrophages by others (C), Microglia by others (D), Monocytes by others (E), Myelinated oligodendrocytes by others (F), Precursor oligodendrocytes by others (G), new oligodendrocytes by others (H) and Neuron by others (I). (J) Only significant overlap of cell type genes identified (A–I) with another cell type are shown. For this figure the following R packages were used: GeneOverlap_1.18.0, VennDiagram_1.6.20, limma_3.38.3, edgeR_3.24.3 and simple comparison expression means [37, 38, 39, 41, 42].

0.2 0.4 0.6 0.8 Color Key

S. Mukherjee



Figure 5. Immune cell type (macrophage, microglia and monocyte) marker genes were enriched in WGCNA modules associated with AD, PD, HD and MS. A: Gene overlap between AD, PD, HD and MS associated WGCNA modules from microarray (ModArray_04) and RNA-seq (ModRNAseq_06 RNA-seq). B: Table showing most significant overlap of cell type signature genes in ModArray_04 microarray WGCNA module and ModRNAseq_06 RNA-seq WGCNA module. C, D: Gene Ontology (GO) Biological Process analysis of ModArray_04 microarray WGCNA module (C) and ModRNAseq_06 RNA-seq WGCNA module (D). For this figure the following R packages were used: EnrichR_1.0, WGCNA_1.66 and ggplot2_3.1.1 [34,43,44,48].

Table 2. Significant cell types enriched in microarray and RNA-seq WGCNA modules associated with AD, PD, HD and MS.

WGCNA_Array_ module_names	DiseaseAD	DiseaseHD	DiseasePD	DiseaseMS	p-value_ Cell_Type_ Enrichment_ in_Array_	Cell_Type_ Marker_Genes	WGCNA_RNA-seq_ module_names	p-value_ Cell_Type_ Enrichment_ in_RNA-
	p-value	p-value	p-value	p-value	WGCNA_ Modules			seq_ WGCNA_ Modules
ModArray_03	9.47E-09	2.89E-13	1.86E-102	3.16E-08	3.26E-16	neuronBYothers	ModRNAseq_14	3.78E-80
ModArray_04	1.92E-36	6.27E-47	0.0005408	4.04E-103	7.77E-08	macrophagesBYothers	ModRNAseq_06	0.00389653
ModArray_04	1.92E-36	6.27E-47	0.0005408	4.04E-103	2.91E-21	microgliaBYothers	ModRNAseq_06	8.32E-39
ModArray_04	1.92E-36	6.27E-47	0.0005408	4.04E-103	7.38E-06	monocytesBYothers	ModRNAseq_06	5.30E-02
ModArray_06	5.23E-05	5.95E-08	9.88E-52	3.21E-81	3.30E-09	myeoligoBYothers	ModRNAseq_08	1.92E-26
ModArray_10	2.17E-42	4.70E-61	0.0216352	9.59E-53	3.85E-17	astroBYothers	ModRNAseq_07	1.50E-142
ModArray_14	1.66E-10	6.97E-10	3.05E-70	2.09E-43	5.43E-30	endoBYothers	ModRNAseq_18	8.54E-85
ModArray_15	1.09E-64	1.50E-84	2.85E-17	1.36E-25	0.00020016	neuronBYothers	ModRNAseq_02	NA
ModArray_20	1.61E-09	9.87E-11	2.28E-12	1.24E-11	1.49E-19	astroBYothers	ModRNAseq_07	1.50E-142

3.5. Gene ontology (GO) annotation of immune cell type enriched AD, PD, HD and MS associated WGCNA modules

Immune system was the most significant biological process gene ontology (GO) in both ModArray_04 microarray WGCNA module and ModRNAseq_06 RNA-seq WGCNA module (Figure 5C,D). Immune process, GO "neutrophil activation involved in immune response" (GO:0002283) was largest in both ModArray_04 (overlap 71/484 p-value 1.43E-42) microarray WGCNA module and ModRNAseq_06 (overlap 83/ 484 p-value 7.66E-08) RNA-seq WGCNA module (Figure 5C,D). Other large GOs in ModArray_04 microarray WGCNA module were related to cell signaling, such as "cytokine-mediated signaling pathway (GO:0019221) overlap 70/634 p-value 3.46E-34" and "positive regulation of intracellular signal transduction (GO:1902533) overlap 36/480 pvalue 5.58E-14" (Figure 5C). Large GOs related to cell signaling were also present in ModRNAseq_06 RNA-seq WGCNA module, such as "protein phosphorylation (GO:0006468) overlap 57/471 p-value 0.000911972" and "toll-like receptor signaling pathway (GO:0002224) overlap 25/87 p-value 8.78E-09" (Figure 5D). These results are consistent with immune cell type enrichment in ModArray_04 microarray WGCNA module and ModRNAseq_06 RNA-seq WGCNA module.

4. Discussion

In MS patients, immune system attacks myelin surrounding neurons and is the most studied central nervous system immune dysfunction [49]. Several recent studies now indicate that a dysfunctional immune system also underlies pathology of other central nervous system diseases, such as neurodegenerative diseases AD, PD and HD [13, 16, 17, 29, 50, 51]. Recently, Mukherjee et al showed that resident brain immune cells called microglia, are a major cell type enriched in gene networks associated with aging and neurodegenerative diseases, AD, PD and HD [52]. Growing evidence suggests that with aging there is an increase in systemic inflammation and collapse of blood brain barrier [19, 20]. However, how immune dysfunction in neurodegenerative diseases compares to MS, a disease known to be caused by immune dysfunction, is not well understood. To address this question, regulatory molecular networks underlying neurodegenerative diseases (AD, PD and HD) and MS were constructed, and immune system characteristics of these networks were evaluated.

To enable comparison of disease effects on gene expression, transcriptome gene expression data were normalized with SVA + LM adjustment to remove effects of age and other covariants on gene expression. A caveat of this work is that association of RNA-seq WGCNA modules with PD was not always significant compared to microarray WGCNA modules. This is probably because RNA-seq PD samples were from cortex while microarray samples were form substantia nigra. However, in spite of this limitation cell type enrichment results were same for significant PD associated microarray WGCNA modules and RNA-seq WGCNA modules (Table2).

Different disease (AD, PD, HD and MS) associated WGCNA modules were enriched by each of the major brain cell type signature genes (neurons, astrocytes, oligodendrocytes and endothelial cells) (Table2). This suggests that all major brain cell type occupy separate AD, PD, HD and MS disease modules and gene networks. Interestingly, all immune cells type signature genes (microglia, monocyte and macrophage) were enriched in the same ModArray_04 microarray WGCNA module and ModRNAseq_06 RNA-seq WGCNA module, associated with AD, PD, HD and MS diseases (Table2, Figure 5A,B). Though macrophages and monocytes shared significant number of common signature genes, microglia signature genes did not significantly overlap with either of them (Figure 4J). This rules out that the three innate immune cell types occur in the same module simply because they have significant number of common signature genes. Taken together, these results suggest that all immune cells underlie a common shared network, which is distinctly different from networks occupied by other brain cell types in AD, PD, HD and MS. One caveat here is that the cell type signature genes and disease modules were obtained from different species, mice and human, respectively.

The AD, PD, HD and MS associated ModArray_04 microarray WGCNA module and ModRNAseq_06 RNA-seq WGCNA module, enriched by all three immune cells (macrophages, microglia and monocytes) was characterized by several immune system related GO terms, "neutrophil activation involved in immune response", "cytokine-mediated signaling pathway", "positive regulation of intracellular signal transduction", "protein phosphorylation" and "toll-like receptor signaling pathway". Taken together, these results suggest a significant role of immune system in the context of neurodegenerative diseases (AD, PD, HD) and MS. In this study, the module or network where immune system is significantly involved in these diseases has been identified. Though more research is required to understand the interaction between immune system and neurodegenerative diseases (AD, PD and HD), one study suggests that microglia are activated by monocytes and macrophages through systemic chronic inflammation in neurodegenerative diseases [53].

Drug targeting strategies involve systemic targeting (specific targeting of organs, tissue and cells) and intracellular targeting (specific targeting of genes and pathways). The basis of network medicine is that diseases occur due to perturbation of entire gene networks and not due to a few individual genes [54]. Several clinical trials targeting the immune system have been undertaken in AD, PD, HD and MS patients. In these trials, anti-inflammatory drugs were found to be ineffective in AD patients, while inhibitors targeting specific inflammatory signaling proteins were found beneficial [55, 56]. A recent study showed that immune-suppressor drugs, inosine monophosphate dehydrogenase inhibitors and corticosteroids, lowered the risk of PD [57]. In this work, several immune pathways were identified in immune cell type enriched in AD, PD, HD and MS associated ModArray_04 microarray WGCNA module and ModRNAseq_06 RNA-seq module. These pathways and WGCNA modules can potentially serve as excellent candidate for intracellular drug targeting in AD, PD, HD and MS treatment. Additionally,

identification of a common underlying immune response in AD, PD, HD and MS diseases in this study, presents a great opportunity to repurpose drugs developed for one of the diseases for another disease. An example of successful drug repurposing, is the approval of a MS treatment drug for immunomodulator targeting macrophages called fumaric acid ester dimethylfumarate (DMF) for HD treatment [58].

In 2015, Durrenberger et al found common immune pathways associated with multiple neurological diseases (AD, PD, HD, ALS, MS and schizophrenia) using microarray datasets [30]. Though this work requires validation with RNA-seq datasets, it raises the exciting possibility that immune cell types (microglia, monocytes and macrophages) associated with neurodegenerative diseases (AD, PD and HD), may also be associated with non-neurodegenerative psychiatric diseases, such as schizophrenia. More research is required to understand the role of innate and adaptive immune system in neurodegenerative diseases, and non-neurodegenerative diseases to identify a safe and effective therapeutic window for immune system drug targeting.

Declarations

Author contribution statement

Shradha Mukherjee: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

For datasets: AD, PD, HD, MS and control microarray (GSE33000, GSE20333, GSE26927, GSE20164, GSE20292, GSE108000 and GSE43490), and RNA-seq (GSE53697, GSE64810, GSE68719, GSE100297 and GSE67333) datasets were obtained from NCBI GEO.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

This work was conducted by the author using free open access resources, personal resources and personal funds from 2017 to 2020. This work is self-owned and not affiliated with any organization/university. The author would like to express deepest gratitude to the journal's publication team for their support. Other open source and free resources utilized in this work are acknowledged below.

For datasets: AD, PD, HD, MS and control microarray (GSE33000, GSE20333, GSE26927, GSE20164, GSE20292, GSE108000 and GSE43490), and RNA-seq (GSE53697, GSE64810, GSE68719, GSE100297 and GSE67333) datasets were obtained from NCBI GEO. Mouse neuron, astrocyte, oligodendrocyte, endothelial cells, microglia, macrophages and monocyte cell type RNA-seq (GSE52564, GSE86573 and SRP035309) datasets were also obtained from NCBI GEO. The author is most grateful to NCBI GEO https://www.ncbi.nlm.nih.gov/geo/ for storing and making these datasets open access. The author also appreciates organizations and investigators who generously submitted these datasets to NCBI GEO and NCBI SRA.

<u>For full text publications access</u>: The author is most grateful to NIH/ NLM (U.S. National Institute of Health's National Library of Medicine) for free access to full-text scientific publications on www.ncbi.nlm.nih .gov/pmc/ PubMed Central.

<u>For computational pipelines</u>: Citations and descriptions of all computational pipelines are given under methods section. Additionally, the author would like to give special thanks to Dr. Steve Horvath, Professor at Department of Human Genetics and Biostatistics, University of California Los Angeles (UCLA) and Dr. Jeffrey T. Leek, Professor at Department of Biostatistics, Johns Hopkins for their publicly available WGCNA/enrichment/Preservation analyses and SVA/Regression analyses computational pipelines, respectively. Dr. Steve Horvath and Dr. Jeffrey T. Leek's computational tutorials were integral to this work.

<u>For computational platform</u>: Computation was performed on personal laptops (MacBook Pro and Windows OS) with high-speed internet connection and computational tools. RNA-seq datasets were mapped using National Science Foundation (NSF) supported open-source Cyverse Discovery Environment.

References

- C. Franceschi, et al., The continuum of aging and age-related diseases: common mechanisms but different rates, Front. Med. 5 (2018) 61.
- [2] A.M. Fjell, et al., What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus, Prog. Neurobiol. 117 (2014) 20–40.
- [3] L.E. Hebert, L.A. Beckett, P.A. Scherr, D.A. Evans, Annual incidence of Alzheimer disease in the United States projected to the years 2000 through 2050, Alzheimer Dis. Assoc. Disord. 15 (2001) 169–173.
- [4] A. Reeve, E. Simcox, D. Turnbull, Ageing and Parkinson's disease: why is advancing age the biggest risk factor? Ageing Res. Rev. 14 (2014) 19–30.
- [5] C. Yapijakis, Huntington disease: genetics, prevention, and therapy approaches, Adv. Exp. Med. Biol. 987 (2017) 55–65.
- [6] M.A. Nance, Genetics of huntington disease, Handb. Clin. Neurol. 144 (2017) 3–14.[7] S.S. Chaganti, E.A. McCusker, C.T. Loy, What do we know about Late Onset
- Huntington's Disease? J. Huntingt. Dis. 6 (2017) 95–103. [8] L.D. Osellame, T.S. Blacker, M.R. Duchen, Cellular and molecular mechanisms of
- (c) LD, oschndrig function, Best Pract. Res. Clin. Endocrinol. Metabol. 26 (2012) 711–723.
- [9] R.H. Swerdlow, J.M. Burns, S.M. Khan, The Alzheimer's disease mitochondrial cascade hypothesis, J. Alzheim. Dis.: JAD 20 (Suppl 2) (2010) S265–279.
- [10] B. Thomas, M.F. Beal, Parkinson's disease, Hum. Mol. Genet. 16 (2) (2007) R183–194.
- [11] S.E. Browne, M.F. Beal, The energetics of Huntington's disease, Neurochem. Res. 29 (2004) 531–546.
- [12] R.A. Floyd, Neuroinflammatory processes are important in neurodegenerative diseases: an hypothesis to explain the increased formation of reactive oxygen and nitrogen species as major factors involved in neurodegenerative disease development, Free Radic. Biol. Med. 26 (1999) 1346–1355.
- [13] S. Mukherjee, C. Klaus, M. Pricop-Jeckstadt, J.A. Miller, F.L. Struebing, A microglial signature directing human aging and neurodegeneration-related gene networks, Front. Neurosci. 13 (2019) 2.
- [14] T. Town, et al., Blocking TGF-beta-Smad2/3 innate immune signaling mitigates Alzheimer-like pathology, Nat. Med. 14 (2008) 681–687.
- [15] T.R. Jay, et al., Disease progression-dependent effects of TREM2 deficiency in a mouse model of Alzheimer's disease, J. Neurosci. : Off. J. Soc. Neurosci. 37 (2017) 637–647.
- [16] H. Keren-Shaul, et al., A unique microglia type Associated with restricting development of Alzheimer's disease, Cell 169 (2017) 1276–1290 e1217.
- [17] D.H. Cribbs, et al., Extensive innate immune gene activation accompanies brain aging, increasing vulnerability to cognitive decline and neurodegeneration: a microarray study, J. Neuroinflammation 9 (2012) 179.
- [18] T.R. Hammond, S.E. Marsh, B. Stevens, Immune signaling in neurodegeneration, Immunity 50 (2019) 955–974.
- [19] A. Montagne, et al., Blood-brain barrier breakdown in the aging human hippocampus, Neuron 85 (2015) 296–302.
- [20] C. Delaney, M. Campbell, The blood brain barrier: insights from development and ageing, Tissue Barriers 5 (2017), e1373897.
- [21] K. Faguy, Multiple sclerosis: an update, Radiol. Technol. 87 (2016) 529–550.
- [22] P.P. Axisa, D.A. Hafler, Multiple sclerosis: genetics, biomarkers, treatments, Curr. Opin. Neurol. 29 (2016) 345–353.
- [23] D. Sadovnick, The place of environmental factors in multiple sclerosis: genes, environment and the interactions thereof in the etiology of multiple sclerosis, Rev. Neurol. (2019).
- [24] S. Fernandez-Menendez, M. Fernandez-Moran, I. Fernandez-Vega, A. Perez-Alvarez, J. Villafani-Echazu, Epstein-Barr virus and multiple sclerosis. From evidence to therapeutic strategies, J. Neurol. Sci. 361 (2016) 213–219.
- [25] C. International Multiple Sclerosis Genetics, et al., Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis, Nat. Genet. 45 (2013) 1353–1360.

S. Mukherjee

- [26] M.M. Wang, D. Miao, X.P. Cao, L. Tan, L. Tan, Innate immune activation in Alzheimer's disease, Ann. Transl. Med. 6 (2018) 177.
- [27] M. Bjorkqvist, et al., A novel pathogenic pathway of immune activation detectable
- before clinical onset in Huntington's disease, J. Exp. Med. 205 (2008) 1869–1877.
 [28] V.H. Perry, Innate inflammation in Parkinson's disease, Cold Spring Harbin Perspect Med. 2 (2012) a009373.
- [29] J.A. Miller, S. Horvath, D.H. Geschwind, Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways, Proc. Natl. Acad. Sci. U. S. A 107 (2010) 12698–12703.
- [30] P.F. Durrenberger, et al., Common mechanisms in neurodegeneration and neuroinflammation: a BrainNet Europe gene expression microarray study, J. Neural. Transm. 122 (2015) 1055–1068.
- [31] S. Mukherjee, R. Brulet, L. Zhang, J. Hsieh, REST regulation of gene networks in adult neural stem cells, Nat. Commun. 7 (2016) 13360.
- [32] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, Bioinformatics 25 (2009) 1105–1111.
- [33] S. Anders, P.T. Pyl, W. Huber, HTSeq-a Python framework to work with highthroughput sequencing data, Bioinformatics 31 (2015) 166–169.
- [34] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016. https://ggplot2.tidyverse.org.
- [35] T. Wei, V. Simko, R Package "corrplot": Visualization of a Correlation Matrix (Version 0.84), 2017. Available from, https://github.com/taiyun/corrplot.
- [36] S. Davis, P.S. Meltzer, GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor, Bioinformatics 23 (2007) 1846–1847.
- [37] M.E. Ritchie, et al., Limma powers differential expression analyses for RNAsequencing and microarray studies, Nucleic Acids Res. 43 (2015) e47.
- [38] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.
- [39] Z. Zhao, et al., Comprehensive RNA-seq transcriptomic profiling in the malignant progression of gliomas, Scient. Data 4 (2017) 170024.
- [40] A. Reiner, D. Yekutieli, Y. Benjamini, Identifying differentially expressed genes using false discovery rate controlling procedures, Bioinformatics 19 (2003) 368–375.
- [41] L. Shen, M. Sinai, GeneOverlap: Test and Visualize Gene Overlaps. R Package Version 1.22.0, 2019. http://shenlab-sinai.github.io/shenlab-sinai/.

- [42] H. Chen, P.C. Boutros, VennDiagram: a package for the generation of highlycustomizable Venn and Euler diagrams in R, BMC Bioinf. 12 (2011) 35.
- [43] P. Langfelder, S. Horvath, Fast R functions for robust correlations and hierarchical clustering, J. Stat. Software 46 (2012).
- [44] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.
- [45] J.T. Leek, J.D. Storey, Capturing heterogeneity in gene expression studies by surrogate variable analysis, PLoS Genet. 3 (2007) 1724–1735.
- [46] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (2012) 882–883.
- [47] J.T. Leek, Svaseq: removing batch effects and other unwanted noise from sequencing data, Nucleic Acids Res. 42 (2014).
- [48] E.Y. Chen, et al., Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool, BMC Bioinf. 14 (2013) 128.
- [49] M. Filippi, et al., Multiple sclerosis, Nat. Rev. Dise. Prim. 4 (2018) 43.
- [50] M.T. Heneka, M.P. Kummer, E. Latz, Innate immune activation in neurodegenerative disease, Nat. Rev. Immunol. 14 (2014) 463–477.
 [51] D.G. Walker, I.F. Lue, Immune phenotypes of microglia in human
- [51] D.G. Walker, L.F. Lue, Immune phenotypes of microglia in human neurodegenerative disease: challenges to detecting microglial polarization in human brains, Alzheimer's Res. Ther. 7 (2015) 56.
- [52] S. Mukherjee, Genomics-Guided Immunotherapy for Precision Medicine in Cancer, Cancer Biother. Radiopharm. 34 (8) (2019).
- [53] Y. Oishi, I. Manabe, Macrophages in age-related chronic inflammatory diseases, NPJ Aging Mech. Dis. 2 (2016) 16018.
- [54] A.L. Barabasi, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease, Nat. Rev. Genet. 12 (2011) 56–68.
- [55] M.J. Daniels, et al., Fenamate NSAIDs inhibit the NLRP3 inflammasome and protect against Alzheimer's disease in rodent models, Nat. Commun. 7 (2016) 12504.
- [56] S. Prokop, K.R. Miller, F.L. Heppner, Microglia actions in Alzheimer's disease, Acta Neuropathol. 126 (2013) 461–477.
- [57] B.A. Racette, et al., Immunosuppressants and risk of Parkinson disease, Ann. Clin. Transl. Neurol. 5 (2018) 870–875.
- [58] G. Ellrichmann, C. Reick, C. Saft, R.A. Linker, The role of the immune system in Huntington's disease, Clin. Dev. Immunol. 2013 (2013) 541259.