

## ARTICLE OPEN



# Interdependence of metals and its binding proteins in Parkinson's disease for diagnosis

Athira Anirudhan<sup>1</sup>, Paramasivam Prabu<sup>2</sup>, Jaya Sanyal<sup>3</sup>, Tapas Kumar Banerjee<sup>4</sup>, Gautam Guha<sup>5</sup>, Ram Murugesan<sup>6</sup> and Shiek S. S. J. Ahmed<sup>6</sup>✉

Metalloproteins utilizes cellular metals which plays a crucial function in brain that linked with neurodegenerative disorders. Parkinson's disease (PD) is a neurodegenerative disorder that affects geriatric population world-wide. Twenty-four metal-binding protein networks were investigated to identify key regulating protein hubs in PD blood and brain. Amongst, aluminum, calcium, copper, iron, and magnesium protein hubs are the key regulators showing the ability to classify PD from control based on thirty-four classification algorithms. Analysis of these five metal protein hubs showed involvement in environmental information processing, immune, neuronal, endocrine, aging, and signal transduction pathways. Furthermore, gene expression of functional protein in each hub showed significant upregulation of EFEMP2, MMP9, B2M, MEAF2A, and TARDBP in PD. Dysregulating hub proteins imprint the metal availability in a biological system. Hence, metal concentration in serum and cerebrospinal fluid were tested, which were altered and showed significant contribution towards gene expression of metal hub proteins along with the previously reported PD markers. In conclusion, analyzing the levels of serum metals along with the gene expression in PD opens up an ideal and feasible diagnostic intervention for PD. Hence, this will be a cost effective and rapid method for the detection of Parkinson's disease.

*npj Parkinson's Disease* (2021)7:3; <https://doi.org/10.1038/s41531-020-00146-7>

## INTRODUCTION

Parkinson disease (PD) is one of the most common neurodegenerative disorders affects elderly populations. Over 10 million people are affected by PD world-wide<sup>1</sup>. The major hallmark of PD is an irreversible loss of dopaminergic neurons at substantia nigra region of the brain. Loss of dopaminergic neurons exhibit cardinal symptoms such as tremor, rigidity, and bradykinesia. Diagnosis of PD relies on medical history and radiological imaging<sup>2,3</sup>. Currently, there is no molecular markers are available to detect PD. Several epidemiological studies suggest multiple risk factors for PD, which includes aging, genetic mutation, and environmental exposure of metals and pesticides<sup>4,5</sup>. The genetic association of PD is well established, which reports PARK loci in most of the GWAS studies<sup>6</sup>. In contrast, the mechanism of metals and pesticide exposures contributing PD is largely unknown.

Emerging metallomic approach helps to investigate the role of metals in a biological system<sup>7</sup>.

Metallomics is the large scale analysis of metals in cells, biofluids, or tissues. It uses high throughput the techniques such as atomic absorption spectroscopy, inductively coupled plasma mass spectrometry (ICP-MS), and inductively coupled optical emission spectrometry for the detecting and estimating the metal concentration in the biological system. Particularly, metallomic analysis of cerebrospinal fluid (CSF), serum, urine, and saliva has proven useful to identify metal biomarkers for several diseases. Additionally, metallomics provide a functional insight on metal interacting molecules such as DNA, RNA, protein, and metabolites. Of metal interacting molecules, metal-binding protein (metalloproteins) cover almost one-third of the human proteome<sup>8</sup>. Metal act as a cofactor that activates protein for basic molecular and

cellular function<sup>9</sup>. Also, metal take part in stabilizing the conformational state of a protein in response to the substrate. Alternatively, metal generate oxidative stress, which serves as a major causative factor for most of the neurological diseases like Alzheimer's, Parkinson's, and Huntington's disease<sup>10,11</sup>. In recent times integrating metallome with other omics data using systems biology has gained a better understanding of metal-mediated pathogenesis in various complex diseases<sup>12,13</sup>.

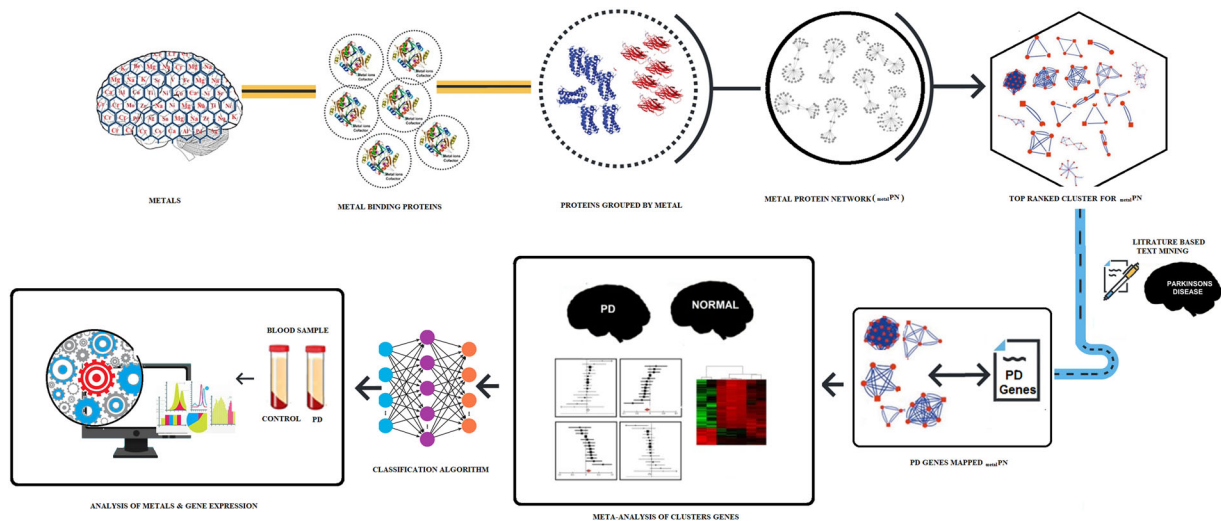
This study implement a systems biological approach (Fig. 1) to dissect the complexity of the metal-binding proteins involved in the pathogenesis of PD. Our approach integrates data mining of metal-binding protein, interactome, and meta-analysis to reveal the significant metal-associated protein hubs in PD. Additionally, we use classification algorithms in detecting highly contributing protein hubs, that classify PD from normal based on gene expression data. Further, the significant metal of each protein hub was measured in blood serum and CSF. Simultaneously, the gene expression of functionally important protein in each hub was analyzed, that reveals the interdependency between metal concentration and gene expression in PD pathogenesis which facilitate the application of metals and its genes as a biomarker in a clinical setting.

## RESULTS

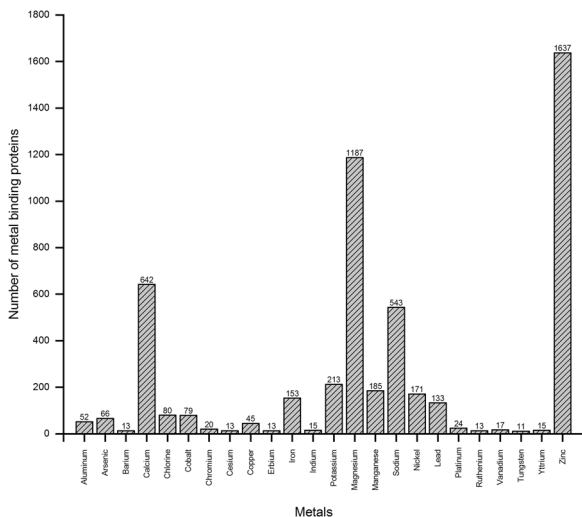
### Data collection and PPI

We implement a systems biological workflow (Fig. 1) to illustrate the mechanism of metal mediating metalloprotein, a causative factor for PD. The metal-binding proteins were collected from various biological databases and curated to have 41 metals that

<sup>1</sup>Faculty of Allied Health Sciences, Chettinad Academy of Research and Education CARE, Kelambakkam 603103, India. <sup>2</sup>School of Medicine, Department of Neurology, University of New Mexico Health Sciences Center, University of New Mexico, New Mexico, USA. <sup>3</sup>Department of Anthropology, University of Delhi, Delhi 110007, India. <sup>4</sup>Department of Neurology, National Neurosciences Centre, Kolkata, India. <sup>5</sup>Department of Neurology, Nil Ratan Sircar Medical College and Hospital, Kolkata, India. <sup>6</sup>Drug Discovery & Omics Lab, Faculty of Allied Health Sciences, Chettinad Academy of Research and Education, Kelambakkam 603103, India. ✉email: shieksjahmed@gmail.com



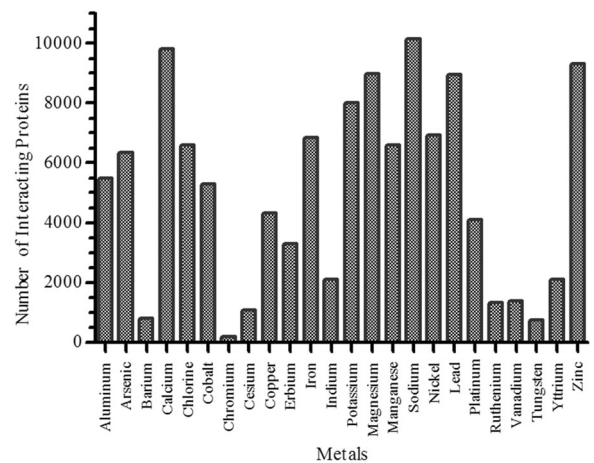
**Fig. 1** Systems biology framework describing the workflow of data collection, integration and analysis of metals and metal-binding proteins in Parkinson's disease.



**Fig. 2** The Bar diagram indicating the number of metal-bound proteins collected for each metal from the protein database.

bounded to 5368 proteins. Among which twenty-four metal-binding proteins groups were selected containing a minimum of ten proteins for each metal. For instance, zinc acts as a cofactor for 1637 proteins, whereas tungsten was noticed with minimum of 11 proteins (Fig. 2). Each metal-binding proteins group was subjected to protein-protein interaction to construct metal protein network ( $_{\text{metal}}\text{PN}$ ). For instance, copper was the cofactor for 45 proteins that form copper  $_{\text{metal}}\text{PN}$  containing 4384 interacting proteins (see Supplementary Fig. 1). Overall, 24  $_{\text{metal}}\text{PN}$ s were constructed, which explicates the complex behavior of the metalloproteins with its interacting proteins (Fig. 3). Of every 24  $_{\text{metal}}\text{PN}$ s, 24 randomized, and 24 molecular-pathway networks were generated and subjected to topological analysis as described under the methodology section. Seven important topological parameters were determined for all 72 networks, which describes their network properties (see Supplementary Table 1).

**Principal component analysis (PCA), clustering, and text mining**  
An unsupervised PCA was performed to check the grouping ability of 24  $_{\text{metal}}\text{PN}$ s with randomized and molecular-pathway networks based on the topological parameters. Two distinct classification



**Fig. 3** The Bar diagram indicate number of interacting protein in each  $_{\text{metal}}\text{PN}$ .

groups were noticed in the PCA model (see Supplementary Fig. 2) group 1—random networks; group 2—molecular pathway networks and  $_{\text{metal}}\text{PN}$ s. The principle component 1 (PC1) accounts for 56.6% and component 2 accounts for 15% of the total variance. Together both the components explain around 71.6% of the data variability. All 24  $_{\text{metal}}\text{PN}$ s were grouped with 24 molecular-pathway networks indicating that the constructed  $_{\text{metal}}\text{PN}$ s are biologically meaningful. Whereas, the 24 randomized networks were distinctly separated in PCA plot (see Supplementary Fig. 2).

Further, the top-ranked protein hub was identified for each  $_{\text{metal}}\text{PN}$  using ClusterONE algorithm. Among 24 over-representing hubs, 18 showed the association with PD containing minimum of one text-mined PD protein. For example, the hub derived from copper  $_{\text{metal}}\text{PN}$  contains six proteins of which all were previously reported in PD confirming the copper association with PD.

#### Meta-analysis

To determine the regulation of selected 18 hubs in PD brain and blood tissue. The gene expression data of brain (GSE7621, GSE8397, GSE19587, GSE20141, GSE20146, GSE20163, GSE20164, GSE20168, GSE20186, GSE20291, GSE20292, GSE20295, GSE28894, and GSE49036) and blood (GSE6613, GSE22491, GSE54536, and GSE72267) were retrieved from NCBI-Geo Dataset and

**Table 1.** Metal concentrations in serum and CSF determined using atomic absorption spectroscopy showing significant change in aluminum, calcium, copper, iron, and magnesium in PD compared to control.

Metal	SERUM (Control = 87; PD = 87)			CSF (Control= 42; PD = 42)		
	Control (µg/L)	PD (µg/L)	<i>P</i> -value	Control (µg/L)	PD (µg/L)	<i>P</i> -value
Aluminum	4.11 ± 1.52	4.755 ± 1.4	<0.02 <sup>a</sup>	3.12 ± 0.886	2.43 ± 1.42	<0.03 <sup>a</sup>
Calcium	64,825 ± 1514	72,303 ± 1720	<0.001 <sup>a</sup>	26,291.15 ± 5411.9	28,447.38 ± 3378.10	<0.03 <sup>a</sup>
Copper	1133 ± 127	909 ± 337	<0.001 <sup>a</sup>	27 ± 4.78	24.4 ± 8.06	<0.029 <sup>a</sup>
Iron	1265 ± 439	1091 ± 434	<0.009 <sup>a</sup>	221 ± 28	172 ± 83.8	<0.002 <sup>a</sup>
Magnesium	20,031 ± 1923	21,276 ± 1915	<0.0006 <sup>a</sup>	23,030 ± 2659	25,466 ± 3512	<0.00055 <sup>a</sup>

<sup>a</sup>Statistical significance (*P*-value ≤ 0.05) was calculated using *t*-test.

EBI-ArrayExpress databases, covering the inclusion criteria of our meta-analysis. The dataset include (normal vs. PD) 291 vs. 304 and 53 vs.104 participants in brain and blood, respectively. The meta-analysis of 302 proteins encoding genes of 18 hubs were analyzed using integrative meta-analysis of gene expression (INMEX), considering the primary goal to identify differentially expressed genes between control vs. PD in brain and blood, respectively. Of 302 genes, 75 were significantly differentially expressed (39 over and 36 were under expressed) in PD brain, which attributed to 16 <sub>metal</sub>PN hubs. Similarly in blood, five genes were upregulated and 86 were downregulated, that attributed to 17 <sub>metal</sub>PN hubs. The genes of these hubs were mapped with their appropriate gene expression data of the brain (normal = 291; PD = 304) and blood (normal = 53; PD = 104) and subjected to the classification algorithm.

#### Classification algorithm based <sub>metal</sub>PN classification

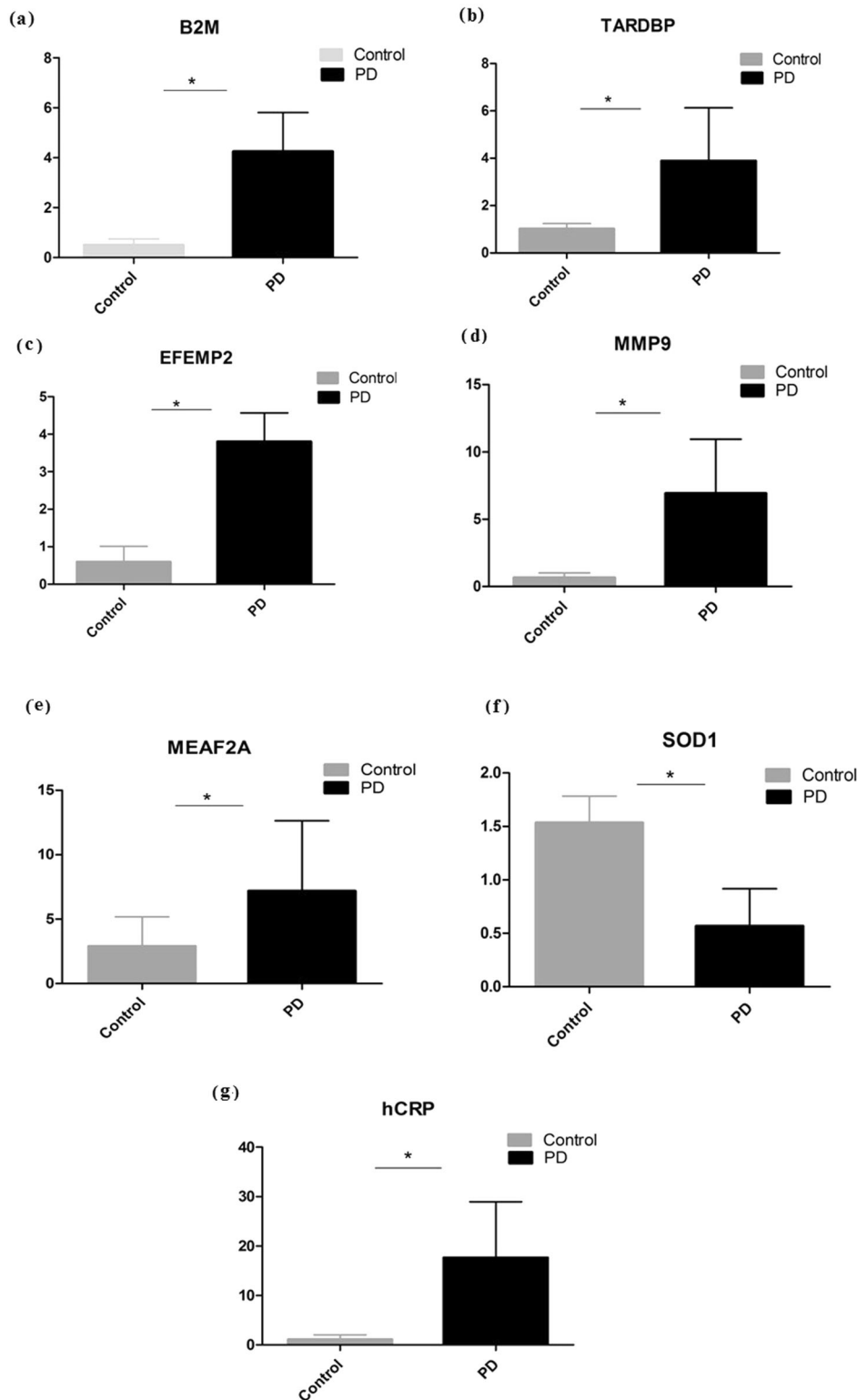
Thirty-four distinct classification algorithms were implemented on each hub with their proteins as features, mapped with corresponding gene expression of brain and blood as instances. For example, the copper hub contains six proteins as features that mapped with appropriate gene expression from the merged brain datasets of 595 instances, normal = 291, PD = 304. Similarly for the blood, the copper hub with the same features contains (53 normal and 104 PD 157 instances). The classification ability of predicting normal or PD by each algorithm for brain and blood was evaluated by accessing the output based on sensitivity, specificity, accuracy and Matthews's correlation coefficient. Based on the performance, top five hubs showing better accuracy between 34 algorithms were selected. For example in brain, iron showed better performance with accuracies ranged between 50 and 91% in all 34 algorithms, whereas cesium hub showed least accuracies between 44 and 56%. Of analyzed hubs, aluminum, calcium, copper, iron, and magnesium were over represented in most of the algorithm and commonly noticed in both brain and blood contributing towards better classification of PD from normal (see Supplementary Fig. 3). Further, the pathway enrichment analysis of 132 proteins of these five hubs showed the association with wide spectrum of molecular pathways related to PD (see Supplementary Fig. 4 and Supplementary Table 2). Particularly, the calcium, copper, iron, and magnesium hubs were commonly regulates 29 molecular pathways (see Supplementary Fig. 5). These pathways attributed to environmental information processing, immune, neuronal, endocrine, aging, and signal transduction mechanism. Subsequent prioritization analysis of proteins in each metal hub suggest EFEMP2, MMP9, B2M, MEAF2A, and TARDBP are functionally important in aluminum, calcium, copper, iron, and magnesium hub, respectively.

#### Metal concentration and gene expression

Estimating these metal concentrations in serum and CSF using atomic absorption spectroscopy (AAS) showed a significant increase in serum aluminum, calcium and magnesium and significant decrease in copper and iron in PD (Table 1). In CSF, significant decrease in aluminum, copper and iron and significant increase in calcium and magnesium were noticed in PD compared to control. The concentrations of analyzed metals were expressed in µg/L with their mean and standard deviation (Table 1). Similarly, the gene expression analysis of EFEMP2, MMP9, B2M, MEAF2A, TARDBP, and hsCRP with qPCR showed significant upregulation, whereas SOD1 was significantly down regulated in PD compared to control (Fig. 4a–g). Further, the interdependency between serum and CSF metal concentrations showed significant change in correlation pattern suggest altered ion transports across blood brain (BB) and blood CSF (BC) barrier in PD (see Supplementary Table 3). Similarly, significant association between metal concentrations with gene expression showed influence of altered serum metals regulates PD hub genes. Interestingly, these metals noticed regulating the gene expression of known oxidative and inflammatory markers (SOD1 and hsCRP) of PD (see Supplementary Tables 4, 5).

#### DISCUSSION

Metal-binding proteins were collected from various biological databases and grouped based on the metal cofactor to have 5368 proteins. For instance, copper acts as a cofactor for 45 proteins (Fig. 2). Of these, α-synuclein is one of the copper-binding proteins considered as a hallmark for PD, localized at presynaptic terminals of neurons. Similar to copper, there are other metals such as aluminum, calcium, chromium, iron, magnesium, lead, and zinc act as a cofactor for several proteins involved in various pathological process of PD. These proteins were grouped based on their bounded metal and subjected to protein-protein interaction to construct metal protein network <sub>metal</sub>PN. Further, the top-ranked hub of each <sub>metal</sub>PN was selected which reveals its core functional mechanism in the network. For example, copper <sub>metal</sub>PN was dissected into six hubs among which the top-ranked hub contains MAP1S, B2M, CD1A, CD1B, CD1D, and CD1E proteins (see Supplementary Fig. 4). Of 24 identified top-ranked hubs, 18 hubs contain minimum of one text-mined PD proteins. In copper hub (see Supplementary Fig. 4), all proteins (MAP1S, B2M, CD1A, CD1B, CD1D, and CD1E) were previously reported in PD<sup>14,15</sup>. Furthermore, the meta-analysis of 18 hubs showed 16 hubs in the brain and 17 in blood were dysregulated in PD. In copper hub, B2M and MAP1S were differentially regulated in both the brain and blood confirming its impairment in PD. B2M is the members of major histocompatibility complex-1 (MHC-1) molecules, widely expressed in various regions of the human brain. Also, it is well established that MHC-1 in neurons play a vital role in synaptic input, synaptic plasticity, and regulates axonal regeneration<sup>16,17</sup>.



**Fig. 4** Quantitative real-time gene expression analysis of hub proteins encoding genes in PBMCs from control vs. PD. **a** B2M, **b** TARDBP, **c** EFEMP2, **d** MMP9, **e** MEAF2A relating copper, magnesium, aluminum, calcium, and iron, respectively. The gene expression of **f** SOD1, and **g** hCRP are the markers for PD. Bars represent the Average, standard deviation (SD) with  $*P$  value  $< 0.05$ .

The expression of MHC-1 is appeared to be influenced in various neuronal pathways of the brain. Further, MHC-1 is involved in cerebellum motor learning and cortical synapse function<sup>16–18</sup>. Several reports suggest altered MHC-1 expression related to neuro-inflammatory mechanism that implicates the immune

mediated neurodegenerative process in diseases, including PD<sup>19,20</sup>. Evidence suggests that B2M has a strong association with complex neurodegenerative diseases such as Alzheimer's and Parkinson's disease<sup>21</sup>. In addition neuronal microtubule-associated protein (MAP1S) in copper hub involved in axonal transport,



regional specialization, and synaptic function<sup>22</sup>. Similar to copper hub, there are other metal hubs noticed to have significant functional association with PD. However, we have tried to explain the importance of our results using the copper hub.

Thirty-four classification algorithms were implemented on each hub to identify the hub showing better classification ability that differentiate PD from normal based on gene expression data. Of analyzed hubs, aluminum, calcium, copper, iron, and magnesium showed better performance in both the tissue for most of the algorithms (see Supplementary Fig. 3). The proteins of these hubs were subjected to enrichment analysis, which showed involvement in immunological, neuronal, endocrinal, and aging process that linked with PD pathogenesis. The proteins of copper hub showed a significant involvement in tight junction, amebiasis and MHC signaling pathways. Tight junctions between the cells contribute fundamental structure of blood-brain barrier (BBB). BBB play a crucial role in maintaining a stable environment of the central nervous system. Altered tight junctions and BBB disruption has been reported in the PD pathogenesis<sup>23</sup>. Overall, our pathway analysis paves a way to understand the role of five significant metal hubs in Parkinson's disease (see Supplementary Table 2).

Dysregulating hub proteins imprint the metal availability in biological system. Thus, metallomic approach was carried out to estimate the concentrations of aluminum, calcium, copper, iron, and magnesium in blood serum and CSF. The analysis of metals showed significant variation in serum and CSF of PD (Table 1) compared to control. All analyzed metals of controls were in reference range of previously published literature<sup>24</sup>. Although, we cannot rule out importance of hub proteins as a possible biomarkers for PD. The hub proteins were prioritized to select functionally important protein from each hub which showed differential regulation in peripheral blood mononuclear cell (PBMC) of PD compared to control (Fig. 4). Mohit Kumar et al.<sup>25</sup>, observed the significant decrease in serum iron levels in PD patients. Similarly, Logroscino et al.<sup>26</sup>, and Hedge et al.<sup>27</sup>, noticed the decreased serum iron concentration and demonstrated its role in PD compared to control. In addition, Jaya Sanyal et al.<sup>24</sup>, showed the decreased copper and iron concentrations in CSF. Similar, results were noticed in this study confirming the importance of these metals in serum and CSF of PD pathology. Also, significant change in serum-CSF metal interdependency of aluminum, calcium, copper, and magnesium between CSF and serum in PD, suggest the impairment of metal transport in BBB and BCB (see Supplementary Table 3).

On the other hand, metal-gene interdependency (see Supplementary Table 4) explores the influence of metals on gene expression in PD. In copper hub, the expression B2M protein encoding gene with the copper concentration showed negative correlation which suggests that the decreased copper concentration increases the B2M gene expression in PD. Increased B2M has been observed in patients with Alzheimer's disease<sup>28</sup>, dementia<sup>29</sup>, Parkinson's diseases<sup>30</sup>, and Schizophrenia patients<sup>31</sup>. Also, Zhang et al. showed B2M induced rat causes cognitive impairments, depression, and anxiety that are closely linked with Parkinson's disease symptoms. Similar metal-gene interdependency was noticed between magnesium-TARDBP, iron-MEAF2A, aluminum-EFEMP2, and calcium-MMP9, suggesting its possible influence in PD. In addition the association between metal with SOD1 and hsCRP marker showed negative correlation which suggest decreased copper concentration increase the expression of SOD1 and hsCRP in PD (see Supplementary Table 5). Also, comparing the interdependency between B2M (copper hub protein) with SOD1 and hsCRP exhibits positive correlation (see Supplementary Table 6). Trist<sup>32</sup> reports that the maturation of SOD1 is blocked by depletion in copper level in brain specific region of PD<sup>33,34</sup>. This lead to decreased antioxidant property of SNc dopaminergic neurons against oxidative stress. Hence, inadequate copper concentration likely lead to the accumulation

of copper deficient wild type SOD1 in brain. Similar results also obtained in our study that supports the decreased copper concentration causes over expression of B2M, SOD1, and hsCRP leading to PD pathogenesis.

Overall our finding provides an insight on the metals and metal-binding proteins in PD pathogenesis. However, our studies have few limitations such as (1) The samples included in this study are south Indian population. Adding the sample from various ethnic background and other neurodegenerative diseases will give more support on reliability of identified metal and gene expression based biomarker for PD. (2) Our study estimated the presence of interdependency between metal and gene expression, having protein expression will provide more information on this association. On the other hand, the strength of our study need to be acknowledged that, (1) This is the first study reporting the metalloprotein network to describe the influence of metal-binding protein in PD. (2) this study integrates metallomic and molecular technique to identify an feasible biomarker from serum. (3) Also, this study provide utilization of classification algorithm, which further reduce complexity of PD diagnosis.

In summary, our study helps to understand the role of metals and its binding protein in Parkinson's disease. The significant changes in CSF and serum level of metals opens up an ideal and feasible diagnostic intervention for PD. Similarly, the interdependency between serum and CSF metal concentration of aluminum, calcium, copper, and magnesium shows significant changes in PD. This alteration makes dysregulation in ion transport across the blood-brain and blood-CSF barrier, which may throw light towards PD diagnosis. Although, the analyzed metals of CSF and serum showed important contribution in PD pathogenesis our potential interest is to identify a feasible markers for diagnosis. Considering the feasibility of both the body fluids, serum holds good for repeated evaluation of disease status and less invasive. Hence, analyzing the levels of serum aluminum, calcium, copper, and magnesium along with the gene expression may be useful in detecting Parkinson's disease.

## METHODS

### Data collection and protein interaction

The proteins containing metal as cofactors were retrieved from the protein databases (<http://www.rcsb.org>, and <http://metalweb.cerm.unifi.it>). Among the collected metal binding proteins, the membrane proteins were excluded due its unbounding of metal at their functional active site involving controlling membrane potential and ionic strength. Using the UniProt mapping tool ([www.uniprot.org](http://www.uniprot.org)), each protein ID was converted into official protein symbol. Further, the collected metal-binding proteins were grouped based on their bounded metal. The protein with multi-metal was assigned to every group of its bounded metal. Twenty-four metal groups were retained containing a minimum of ten proteins in each group. The metal protein network ( $_{metal}PN$ ) was constructed using Cytoscape software for each metal group by extending each metal-binding protein to immediate neighbor interacting protein.

### Principle component analysis based on network properties

To determine the biological significance of 24  $_{metal}PN$ s, the 24 randomized, and 24 molecular pathway networks were generated. Each randomized network was constructed using Erdos-Renyi model. Twenty-four randomized networks were created containing an equal number of nodes corresponding to every 24  $_{metal}PN$ s. Similarly, the 24 molecular pathway networks were constructed by collecting proteins from the 24 randomly selected molecular pathways available in the KEGG database. Seven topological parameters were determined for each network using Cytoscape network analyzer. Principle component analysis (PCA) was performed using R-program, on the variables describing the topological parameters of 72 networks to group them based on their network properties<sup>35</sup>.

## Clustering, text-mining, and meta-analysis

ClusterONE algorithm was implemented on each  $_{\text{metal}}\text{PN}$  to decompose large network into dense protein hubs<sup>36</sup>. Each protein hub represents the core functional component of  $_{\text{metal}}\text{PN}$ . The top-ranked hub of each  $_{\text{metal}}\text{PN}$  was selected to determine its involvement in PD. For which, the text-mining approach was carried out to collect the proteins reported in PD from literature and mapped to each top-ranked hub of 24  $_{\text{metal}}\text{PN}$  to select PD hubs.

## Text-mining

Indigenous R-code was developed to retrieve the abstract for the key terms related to Parkinson's disease in conjunction with "Homo sapiens" from the PubMed database, published between the years 1990–2017. The collected 84,214 abstracts were examined for the presence of 22,853 protein symbols from the dictionary constructed by integrating various proteome databases<sup>37–39</sup>. The R-code looks for the co-occurrence of the terms "Parkinson's disease" and protein symbol. We weigh the associations of protein with the PD based on the occurrence of each protein symbol and PD in the abstract based on the point-wise mutual information method. Further, the frequencies of the collected PD proteins across the abstracts were evaluated to have its presence in a minimum of three abstracts. The identified PD proteins were mapped to the top-ranked hub of each metal protein network ( $_{\text{metal}}\text{PN}$ ) to select PD associated hub.

## Meta-analysis

To determine the regulation of selected hub in PD, the expression of proteins encoding genes in the each hub was subjected to the meta-analysis. For which, the microarray gene expression datasets were extracted by two independent reviewers (AA & JS) from NCBI Gene Expression Omnibus GEO (<http://www.ncbi.nlm.nih.gov/geo/>) and Arrayexpress (<https://www.ebi.ac.uk/arrayexpress/>) using the keyword related to Parkinson's disease. The inclusion criteria were set for the selection of dataset include (1) human case-control study, (2) dataset comparing PD with other neurological disease and healthy control, (3) study conducted in brain and/or blood tissue, (4) availability of raw or processed dataset. Also, we exclude the expression dataset of (1) non-human study, (2) in vivo study, (3) integrated and secondary analysis of expression data, and (4) the microarray dataset with less than two replicate samples per type. Any discrepancy in collected dataset was resolved by the third reviewer (SSJ), while the group discussion. The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines published in 2015 was followed in the selection of datasets for this meta-analysis<sup>40</sup>.

## Meta-analysis of genes in $_{\text{metal}}\text{PN}$ hubs

The collected microarray datasets were subjected to INMEX<sup>41</sup>, and the analysis was carried out by following the procedure of Jose et al.<sup>42</sup>. In brief, each dataset was annotated by converting probe ID to Entrez ID. The gene intensity was log-transformed and limma processed to determine the significant genes  $P$ -value ( $\leq 0.05$ ) dataset between control and PD. For meta-analysis, the datasets were merged to apply combat to decrease the batch effect. Further, Fisher's statistical approach was used to combine significant  $P$ -values of each dataset to determine the significant genes from the merged datasets. For the collected 14 brain datasets, the meta-analysis was conducted to determine the significant genes in PD with  $P$ -value  $\leq 0.05$ . Similarly, the significant PD genes in blood were identified from the merged four datasets.

## Classification algorithm to classify $_{\text{metal}}\text{PN}$ s hubs

Data classification was performed using Weka software (Waikato Environment for Knowledge Analysis)<sup>43</sup>. Thirty-four distinct classification algorithms were used to classify each hub into PD from normal based on the merged expression data of the meta-analysis. The protein of each hub was considered as features with its corresponding gene expression value as input for classification algorithm. The performance was evaluated based on the ten-times, ten-fold cross-validation technique. The threshold independent parameters such as sensitivity, specificity, accuracy and Matthews's correlation coefficient were calculated for each hub. Further, top five hubs showing high classification accuracy between the 34 algorithms were selected and their proteins were subjected to KEGG pathway enrichment analysis<sup>44</sup>. For validation, the hub relating metal was analyzed using AAS in serum of control and PD. Simultaneously, the proteins in each hub was functionally prioritized using ToppGene tool<sup>45</sup>. In ToppGene tool, the text-

mined 2088 PD proteins excluding hub proteins were inputted as training-set and the proteins of each hub was considered as test-set, to identify the functionally importance protein in each hub. Further, the change in expression of these proteins encoding genes were tested in peripheral blood mononuclear cell (PBMC) using Real time qPCR along with the previously known antioxidant (Superoxide dismutase; SOD1) and inflammatory markers (C-reactive protein; hsCRP) of PD.

## Participant selection

The participants were recruited at the National Neuroscience Center (NNC) and Nil Ratan Sircar Medical College and Hospital (NRS), Kolkata, India. The protocol of this study, including sample collection, processing and analysis was approved by the Research Ethics Boards Committee of the participating centers (NNC: Ref.No.Ath/2013) and NRS: Ref.No.25062012). The written informed consent was obtained from all participants prior to sample collection. The PD patients were enrolled and diagnosed by the neurologist based on neurological examination and medical history. To have a true positive PD patients, Unified Parkinson's Disease Rating Scale (UPDRS) and Hoehn and Yahr scale were followed. For the comparative analysis, the participants free from neurological or neuropsychiatric disorders were taken and underwent neurological examinations similar as the patient for confirmation. Additionally, the selected controls are equally matched with the PD participants based on their age, gender, comorbidities such as diabetes, hypertension, and cardio vascular disease (Supplementary Table 8).

## Inclusion and exclusion criteria

Participants for this study were selected by following the inclusion and exclusion criteria. Inclusion criteria: (1) ability to comply with study procedures. (2) participants aged 45–65 years. (3) body mass index (BMI) of 18–35 kg/m<sup>2</sup>. For participants with PD: (1) drug naïve, (2) presence of cardinal symptoms, (3) absence of secondary Parkinsonism because of drugs, (4) no features of prominent oculomotor palsy, cerebellar signs, vocal cord paresis, pyramidal signs, severe orthostatic hypotension, amyotrophy, or apraxia. The exclusion criteria include: (1) history of alcohol abuse or consumption, (2) smoker or tobacco user, (3) previous history of severe systemic diseases, (4) intake of minerals and chelating agents, and (5) acute infections, traumata or surgery in the last one year.

## Sample collection and processing procedure

The CSF and blood were drawn from the same patients and compared with healthy controls. The demographic and clinical characteristics of participants were recorded (see Supplementary Table 8). The CSF (control = 42; PD = 42) and peripheral blood (control = 87; PD = 87) samples were obtained from the participating centers. A few did not show willingness to provide CSF samples. Hence, CSF obtained only from 42 control and 42 PD by following the BioMS-eu consortium guidelines. Ten ml of CSF was obtained by lumbar spinal tap using a traumatic needle. The CSF sample was then centrifuged at 1200 rpm for 10 min at room temperature to collect cell-free supernatant. Similarly, blood (10 + 3 mL) was collected, 10 mL were centrifuged for 10 min at 3000 rpm to separate serum and the remaining 3 mL was used for PBMC isolation. The separated samples were stored at  $-80^{\circ}\text{C}$  for further analysis.

## Analysis of hub associated metals in serum and CSF

All precautions were taken following the NCCLS Guidelines to avoid contaminants while sample collection and processing for the element analysis. Nitric acid based microwave digestion was carried out<sup>46</sup>. For calibration, NIST SRM 3100 series single-element standard solution was used at various concentrations to have standard graphs. Also, the limits of detection (LoDs) for the metals were determined by analyzing blank solutions. The element concentration of aluminum and iron in serum was analyzed using electrothermal (AAS SHIMADZU AA-6200, Kyoto, Japan). Whereas, calcium, copper, and magnesium were analyzed using flame AAS VARIAN AA-240, Varian Inc, (USA). The internal standards were run between every six samples to determine the accuracy and quality of the analysis.

## Gene expression analysis of clusters proteins, antioxidant, and inflammatory marker

Peripheral blood mononuclear cell were isolated using Histopaque-1077 (Sigma-Aldrich) according to the manufacturer's protocol. Further, the TRIzol (Invitrogen) reagent was used for isolation of total RNA from PBMC. The isolated RNA was quantified using Nanodrop 2000 (Thermo Scientific) to construct cDNA from total RNA<sup>47</sup>. Quantitative real-time PCR was performed to determine the expression of protein encoding gene in each metal hub and antioxidant and inflammatory marker between PD and control. The expression analysis was executed using ABI-7000 (Applied Biosystems) with SYBR green master mix (Takara) and gene specific primers (see Supplementary Table 7). The expression of targeted gene was normalized with  $\beta$ -actin housekeeping gene and relative expression was calculated following the  $2^{-\Delta\Delta Ct}$  method.

## Statistical analysis

The statistical analysis was performed using SPSS software. The student *t*-test was performed to determine the statistical difference ( $P \leq 0.05$ ) between the control and PD for the measured metals and PBMC gene expression, respectively. Further, correlation analysis was carried out to determine the interdependency of (1) Metal concentration between serum and CSF, (2) PBMC gene expression with serum metal concentration, (3) PBMC SOD expression with metal and metal hub protein, and (4) PBMC hsCRP expression with metal and metal hub protein.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## DATA AVAILABILITY

The raw gene expression data used in this study were obtained from the NCBI-GEO DataSets (<https://www.ncbi.nlm.nih.gov/gds>). The gene expression dataset can be accessed using accession IDs GSE7621, GSE8397, GSE19587, GSE20141, GSE20146, GSE20163, GSE20164, GSE20168, GSE20186, GSE20291, GSE20292, GSE20295, GSE28894, and GSE49036 for brain. Similarly for blood, GSE6613, GSE22491, GSE54536, and GSE72267 can be used.

## CODE AVAILABILITY

IBM SPSS package version 24.0 was used for the statistical analysis. The meta-analysis based gene expression was executed using integrative meta-analysis of gene expression (INMEX). The indigenous R-code was used for text-mining will be provided on request.

Received: 1 May 2020; Accepted: 12 November 2020;

Published online: 04 January 2021

## REFERENCES

- Surathi, P., Jhunjhunwala, K., Yadav, R. & Pal, P. K. Research in Parkinson's disease in India: a review. *Ann. Indian Acad. Neurol.* **19**, 9–20 (2016).
- Saeed, U. et al. Imaging biomarkers in Parkinson's disease and Parkinsonian syndromes: current and emerging concepts. *Transl. Neurodegener.* **6**, 8 (2017).
- Pagano, G., Niccolini, F. & Politis, M. Imaging in Parkinson's disease. *Clin. Med.* **16**, 371–375 (2016).
- Horowitz, M. P. & Greenamyre, J. T. Gene–Environment interactions in Parkinson's disease: the importance of animal modeling. *Clin. Pharmacol. Ther.* **88**, 467–474 (2010).
- Nandipati, S. & Litvan, I. Environmental exposures and Parkinson's disease. *Int. J. Environ. Res. Public Health* **13**, 881 (2016).
- Tan, E. K. et al. Analysis of GWAS-linked loci in Parkinson disease reaffirms PARK16 as a susceptibility locus. *Neurology* **75**, 508–512 (2010).
- Ishihara, K. & Akiba, S. A comprehensive diverse 'omics' approach to better understanding the molecular pathomechanisms of down syndrome. *Brain Sci.* **7**, 44 (2017).
- Harding, M. M., Nowicki, M. W. & Walkinshaw, M. D. Metals in protein structures: a review of their principal features. *Crystallogr. Rev.* **16**, 247–302 (2010).
- Valdez, C. E., Quentin, A. S., Nechay, M. R. & Alexandrova, A. N. Mysteries of metals in metalloenzymes. *Acc. Chem. Res.* **47**, 3110–3117 (2014).

- Anthony, R. W., Michael, A., Lucio, G. C. & Ashley, I.B. *Biometals in Neurodegenerative Diseases Mechanisms and Therapeutics* (Elsevier Academic Press, USA, 2017).
- Brown, D. R. Metalloproteins and neuronal death. *Metallomics* **2**, 186–194 (2010).
- Maret, W. The metals in the biological periodic system of the elements: concepts and conjectures. *Int. J. Mol. Sci.* **17**, 66 (2016).
- Gómez-Ariza, L., José, T. G. -B., Lorenzo, F. & Arias, A. Analytical characterization of bioactive metal species in the cellular domain (metallomics) to simplify environmental and biological proteomics. *Int. J. Environ. Anal. Chem.* **85**, 255–266 (2005).
- Smith, L. K. et al.  $\beta$ 2-microglobulin is a systemic pro-aging factor that impairs cognitive function and neurogenesis. *Nat. Med.* **21**, 932 (2015).
- Wang, Z. T. et al. Alzheimer's disease neuroimaging initiative. Genome-wide association study identifies CD1A associated with rate of increase in plasma neurofilament light in non-demented elders. *Ageing* **1113**, 4521 (2019).
- Shatz, C. J. MHC class I: an unexpected role in neuronal plasticity. *Neuron* **64**, 40–45 (2009).
- Cebrián, C., Loike, J. D. & Sulzer, D. Neuronal MHC-I expression and its implications in synaptic function, axonal regeneration and Parkinson's and other brain diseases. *Front. Neuroanat.* **8**, 114 (2014).
- McConnell, M. J., Huang, Y. H., Datwani, A. & Shatz, C. J. H2-Kb and H2-Db regulate cerebellar long-term depression and limit motor learning. *Proc. Natl Acad. Sci.* **106**, 6784–6789 (2009).
- Tansey, M. G. & Goldberg, M. S. Neuroinflammation in Parkinson's disease: its role in neuronal death and implications for therapeutic intervention. *Neurobiol. Dis.* **37**, 510–518 (2010).
- Cebrián, C. et al. MHC-I expression renders catecholaminergic neurons susceptible to T-cell-mediated degeneration. *Nat. Commun.* **5**, 3633 (2014).
- Gupta, V. B. et al. Altered levels of blood proteins in Alzheimer's disease longitudinal study: results from Australian Imaging Biomarkers Lifestyle Study of Ageing cohort. *Alzheimer's Dement.* **8**, 60–72 (2017).
- Yang, M. et al. The role of microtubule-associated protein 1B in axonal growth and neuronal migration in the central nervous system. *Neural Regen. Res.* **7**, 842 (2012).
- Kortekaas, R. et al. Blood–brain barrier dysfunction in parkinsonian midbrain in vivo. *Ann. Neurol.* **57**, 176–179 (2005).
- Sanyal, J. et al. Metallomic biomarkers in cerebrospinal fluid and serum in patients with Parkinson's disease in Indian population. *Sci. Rep.* **6**, 35097 (2016).
- Gangania, M. K., Batra, J., Kushwaha, S. & Agarwal, R. Role of iron and copper in the pathogenesis of Parkinson's disease. *Indian J. Clin. Biochem.* **32**, 353–356 (2017).
- Logroscino, G. et al. Altered systemic iron metabolism in Parkinson's disease. *Neurology* **49**, 714–717 (1997).
- Hegde, M. L. et al. Serum trace element levels and the complexity of inter-element relations in patients with Parkinson's disease. *J. Trace Elem. Med. Biol.* **18**, 163–171 (2004).
- Carrette, O. et al. A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease. *Proteomics* **3**, 1486–1494 (2003).
- Zhang, P. et al.  $\beta$ 2-microglobulin induces depressive- and anxiety-like behaviors in rat. *PLoS ONE* **13**, e0198027 (2018).
- Zhang, J. et al. CSF multianalyte profile distinguishes Alzheimer and Parkinson diseases. *Am. J. Clin. Pathol.* **129**, 526–529 (2008).
- Xie, J., Wang, Y., Freeman, M. E. III, Barlogie, B. & Yi, Q.  $\beta$ 2-Microglobulin as a negative regulator of the immune system: high concentrations of the protein inhibit in vitro generation of functional dendritic cells. *Blood J. Am. Soc. Hematol.* **10110**, 4005–4012 (2003).
- Trist, B. G., Hare, D. J. & Double, K. L. A proposed mechanism for neurodegeneration in movement disorders characterized by metal dyshomeostasis and oxidative stress. *Cell Chem. Biol.* **25**, 807–816 (2018).
- Davies, K. M. et al. Copper pathology in vulnerable brain regions in Parkinson's disease. *Neurobiol. Aging* **35**, 858–866 (2014).
- Genoud, S. et al. Subcellular compartmentalisation of copper, iron, manganese, and zinc in the Parkinson's disease brain. *Metallomics* **9**, 1447–1455 (2017).
- David, C. C. & Jacobs, D. J. Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol. Biol.* **1084**, 193–226 (2014).
- Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods* **9**, 471 (2012).
- Muthusamy, B., Thomas, J., Prasad, T. S. K. & Pandey, A. Access guide to human protein pedia. *Curr. Protoc. Bioinform.* **41**, 1–21 (2013).
- Kim, M. S. et al. A draft map of the human proteome. *Nature* **509**, 575 (2014).
- Thul, P. J. & Lindskog, C. The human protein atlas: a spatial map of the human proteome. *Protein Sci.* **27**, 233–244 (2018).
- Moher, D. Preferred reporting items for systematic review and meta-analysis protocols PRISMA-P) 2015 statement. *Syst. Rev.* **41**, 1 (2015).

41. Xia, J. et al. INMEX—a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Res.* **41**, W63–W70 (2013).
42. Jose, A. & Potashkin, J. A. Network-based meta analysis identifies HNF4A and PTBP1 as longitudinally dynamic biomarkers for Parkinson's disease. *Proc. Natl Acad. Sci.* **17**, 2257–2262 (2015).
43. Sriram, T. V. S., VenkateswaraRao, M., SatyaNarayana, G. V., Kaladhar, D. S. V. G. K. & PanduRanga Vital, T. Intelligent Parkinson disease prediction using machine learning algorithms. *Int. J. Eng. Innov. Technol.* **2**, 44–52 (2013).
44. Ogata, H. et al. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
45. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
46. Bader, N. R. Sample preparation for flame atomic absorption spectroscopy: an overview. *Rasayan J. Chem.* **4**, 49–55 (2011).
47. Sun, L. et al. Neuroprotection by IFN- $\gamma$  via astrocyte-secreted IL-6 in acute neuroinflammation. *Oncotarget* **20**, 825:40065 (2017).

## ACKNOWLEDGEMENTS

The authors are thankful to the participants for voluntarily donating their blood for the study. Also, we thank the collaborating institutes for their support throughout the study.

## AUTHOR CONTRIBUTIONS

A.A.: executed the bioinformatics protocol, data analysis, interpretation and written the base manuscript; J.S.: involved in metallomic analysis; T.K.B. and G.G.: screen the participants; P.P.: over viewed the gene expression study; R.M.: Give critical suggestion on study concept, design, statistical analysis, and interpret the data; S.S.J.: study concept, design, pooled the data, interpret the data, finalized, and approve the manuscript. All authors read and approved the final manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41531-020-00146-7>.

**Correspondence** and requests for materials should be addressed to S.S.S.J.A.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021