

RESEARCH ARTICLE

Open Access



# Whole blood gene expression and white matter Hyperintensities

Honghuang Lin<sup>1,2\*</sup> , Claudia Satizabal<sup>1,3</sup>, Zhijun Xie<sup>4</sup>, Qiong Yang<sup>5</sup>, Tianxiao Huan<sup>1,6</sup>, Roby Joehanes<sup>7,8</sup>, Chengping Wen<sup>4</sup>, Peter J. Munson<sup>7</sup>, Alexa Beiser<sup>1,3,5</sup>, Daniel Levy<sup>1,6</sup> and Sudha Seshadri<sup>1,3\*</sup>

## Abstract

**Background:** White matter hyperintensities (WMH) are an important biomarker of cumulative vascular brain injury and have been associated with cognitive decline and an increased risk of dementia, stroke, depression, and gait impairments. The pathogenesis of white matter lesions however, remains uncertain. The characterization of gene expression profiles associated with WMH might help uncover molecular mechanisms underlying WMH.

**Methods:** We performed a transcriptome-wide association study of gene expression profiles with WMH in 3248 participants from the Framingham Heart Study using the Affymetrix Human Exon 1.0 ST Array.

**Results:** We identified 13 genes that were significantly associated with WMH (FDR < 0.05) after adjusting for age, sex and blood cell components. Many of these genes are involved in inflammation-related pathways.

**Conclusion:** Thirteen genes were significantly associated with WMH. Our study confirms the hypothesis that inflammation might be an important factor contributing to white matter lesions. Future work is needed to explore if these gene products might serve as potential therapeutic targets.

**Keywords:** White matter hyperintensities, Gene expression, Epidemiology

## Background

Magnetic resonance imaging (MRI) has been frequently used to identify abnormalities of white matter, detectable as white matter hyperintensities (WMH) [1, 2]. The prevalence of these white matter lesions increases with aging [3–5], and they typically happen long before the onset of clinically manifest neurological conditions [6]. WMH are observed in both apparently healthy individuals [2, 7, 8] and in individuals with stroke and dementia [9]. They have been associated with the risk of a variety of neurological diseases and other adverse outcomes, such as dementia [10–12], cognitive dysfunction [5, 9, 13–15], cardiovascular diseases [7], stroke [16–19], and mortality [20, 21]. The progression of WMH has also been used as a biomarker for predicting outcomes following stroke [22] and as a surrogate endpoint for clinical trials of cerebral small-vessel disease [23].

The pathophysiology of WMH remains poorly understood. Cardiovascular disease risk factors such as hypertension and diabetes are also risk factors for WMH, and small vessel arteriosclerotic disease is thought to be a key mechanism leading to WMH. In recent years, increasing evidence has suggested that chronic inflammation and glial proliferation might also be involved in the pathogenesis of WMH [24]. Moreover, it has been shown that WMH is highly heritable, with heritability estimates ranging from 55 to 73% [25–27]. Several genetic loci have been identified to be associated with WMH [28–31], however, they explain only a small portion of the observed heritability.

Gene expression has proved to be an important intermediate phenotype that helps to bridge genetic variations with many phenotypic traits [32–34]. Xu et al. [35] examined the association of gene expression in blood with WMH. Twenty participants with extensive WMH and 18 participants with minimum WMH were enrolled in the study. A total of 241 genes were found to be differentially expressed ( $P < 0.005$  and at

\* Correspondence: hmlin@bu.edu; susheshad@bu.edu

<sup>1</sup>National Heart Lung and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA, USA

Full list of author information is available at the end of the article



least 1.2 fold difference), many of which are involved in inflammation, oxidative stress, detoxification and hormonal responses. Another study by Simpson et al. [36] examined the association of gene expression in postmortem central nervous system tissue with WMH. Seven participants with WM lesions and 7 participants without lesions were enrolled in the study, and 502 genes were found to be differentially expressed, including genes involved in immunity and ion transport.

These studies have demonstrated that differential gene expression was associated with WMH. However, they examined small samples of highly selected participants, which limits their generalizability. The objective of this study is to assess the association of gene expression with WMH in over 3000 participants from the Framingham Heart Study (FHS).

## Methods

### Study samples

The FHS is a community-based, prospective study initiated in 1948 that aimed to investigate cardiovascular disease and its risk factors in the community. Three generations of participants have been enrolled, and multiple examinations have been performed with an interval of 2 to 8 years [37–39]. At each clinical examination, participants go through extensive physical examination, lifestyle and medical history interview. This analysis is restricted to the second generation (Offspring) of participants who attended the eighth Offspring examination (2005–2008), and the Third Generation participants who attended the second examination (2008–2011). All participants gave written informed consent, and the study was approved by the institutional review board at the Boston University Medical Center.

### WMH measurement

The details for the MRI protocol in FHS have been described previously [9, 40]. In brief, MRI was performed on a 1.5 Tesla Siemens Avanto scanner. Fluid attenuated inversion recovery (FLAIR) sequences were used for the measurement of WMH. The segmentation and quantification of WMH was performed using a semi-automated procedure [41], which has shown high inter-rater reliability [42]. Total intracranial volume (TCV), based on FLAIR, was quantified using the Quanta 2 software package [41].

### RNA extraction and gene expression profiling

Total RNA was isolated from frozen PAXgene blood tubes (PreAnalytiX, Hombrechtikon, Switzerland) and amplified using the WT-Ovation Pico RNA Amplification System (NuGEN, San Carlos, CA) according to the manufacturers'

standard operating procedures. The obtained cDNA was hybridized to the Affymetrix Human Exon 1.0 ST Array (Affymetrix, Inc., Santa Clara, CA). The raw data were quantile-normalized and log<sub>2</sub> transformed, followed by summarization using Robust Multi-array Average [43]. The gene annotations were obtained from Affymetrix NetAffx Analysis Center (version 31). We excluded transcript clusters that were not mapped to RefSeq transcripts, resulting in 17,873 distinct transcripts (17,324 distinct genes) for downstream analysis.

Given that the gene expression was measured from whole blood, the proportion of different cell types might affect gene expression. However, only 2181 participants from the Third Generation cohort had measured cell counts, of which 1225 were included in the current study. For the remaining participants, we used the partial least square method to estimate the cell counts from those with measured cell counts based on the gene expression data [44]. The percentages of each imputed cell type were then normalized, where the negative predicted values were set to 0 and the sum of the percentages for all cell types was set as 100%. Cross-validated estimates of prediction accuracy (R<sup>2</sup>) were 0.61, 0.41, 0.25, 0.83, 0.83, 0.81, 0.89, and 0.25, for white blood cell counts, red blood cell counts, platelet counts, neutrophil percent, lymphocyte percent, monocyte percent, eosinophil percent, and basophil percent, respectively.

### Statistical analyses

WMH measures were log transformed to reduce the skewness of its distribution. Linear mixed effects models were used to test the association between gene expression and WMH volumes, treating the expression of each gene as the dependent measure, and the log-transformed WMH measure as the exposure. The analyses were adjusted for age, sex, and TCV. We also adjusted for the differential cell counts using a fixed effect factor, and for familial relatedness by implementing a random variance-covariance matrix.

In our secondary analyses, we additionally adjusted for smoking, body mass index, systolic blood pressure, diastolic blood pressure, hypertension treatment, total cholesterol, HDL cholesterol, and triglyceride.

In order to correct for multiple testing, we used false discovery rate (FDR) [45], which is defined as the number of incorrectly rejected hypotheses divided by the total number of rejected hypotheses. Significant associations were defined as those with FDR < 0.05. All the analyses were performed using the "lmekin" R package ([www.r-project.org/](http://www.r-project.org/)).

### Overlap with GWAS loci

The summary statistics of GWAS association was obtained from a meta-analysis of participants from multiple

ancestries [46]. Four genetic loci were significantly associated with WMH. At each locus, we obtained all SNPs with  $P$ -value less than  $5 \times 10^{-8}$  (defined as GWAS SNPs). We then examined if any of GWAS SNPs was associated with gene expression in blood using FHS expression quantitative trait loci (eQTL) database [44]. An eGene was defined if its expression was associated with at least one of GWAS SNPs (FDR < 0.05). The association of each of the eGenes with WMH was then examined as described in the previous section.

We also queried the GWAS catalog [47] and extracted variants significantly associated with stroke or dementia ( $P < 5 \times 10^{-8}$ ). For each of these variants, we searched its eGene, and examined the association of eGenes as described in the previous section.

### Construction of gene interaction subnetwork associated with WMH

A dense module searching strategy [48] was used to identify modules enriched with WMH-related genes. The experimentally validated interactions between genes were obtained from the PINA database [49]. Before the searching, each gene was assigned a score to represent its association with WMH. The module searching started with a seed gene that was significantly associated with WMH (FDR < 0.05). Neighboring genes were then added sequentially to the module if the addition increased the overall module score [50], which was defined as  $Z_m = \frac{\sum g_i}{\sqrt{k}}$ , where  $k$  is the number of genes in the module, and  $g_i$  is the score of the gene  $i$ . The searching stopped if no more genes could be added.

## Results

The current study includes 1397 eligible participants from the Offspring Cohort (mean age  $66.4 \pm 9.0$  years, 54.2% women) and 1851 participants from the Third Generation Cohort (mean age  $48.0 \pm 8.5$  years, 54.0% women) who had both gene expression and WMH measured. The descriptive characteristics of the participants are provided in Table 1.

### Association of Gene Expression with WMH volume

As shown in Table 2, a total of 13 genes were significantly associated with WMH (FDR < 5%). Six of them were upregulated, and the remaining seven genes were down-regulated. Figure 1 is the volcano plot showing the association between each gene with WMH. The most significant gene was *IL4R* ( $P = 1.5 \times 10^{-8}$ ), which encodes the alpha chain of the interleukin 4 receptor. The result was similar after excluding participants with stroke, dementia and vascular diseases (Additional file 1: Table S1).

**Table 1** Baseline characteristics of the study participants

Characteristics	Offspring cohort ( $n = 1397$ )	Third generation cohort ( $n = 1851$ )
Women, $n$ (%)	757 (54.2%)	1000 (54.0%)
Age, years (mean $\pm$ SD)	$66.4 \pm 9.0$	$48.0 \pm 8.5$
WMH, median (25th, 75th percentile)	2.41 (1.25, 4.69)	1.28 (0.92, 1.75)
Ln(WMH + 1), median (25th, 75th percentile)	1.23 (0.81, 1.74)	0.82 (0.65, 1.01)
Total cranial volume, median (25th, 75th percentile)	1222 (1143, 1307)	1254 (1170, 1344)
Current smoker, $n$ (%)	97 (6.9%)	174 (9.4%)
Hypertension, $n$ (%)	828 (59.3%)	135 (7.3%)
Body mass index (BMI), median (25th, 75th percentile)	27.52 (24.52, 30.78)	27.11 (23.98, 30.72)
Total cholesterol, median (25th, 75th percentile)	185 (162, 210)	184 (163, 207)
HDL, median (25th, 75th percentile)	56 (45, 68)	57 (47, 71)

In the secondary analysis, we adjusted the model for additional clinical factors (see Methods). As shown in Table 2, most of top hits were still significant, although the association were slightly attenuated.

In order to assess potential effects of imputed cell counts, we performed a sensitivity analysis by including only individuals with measured cell counts, and compared the association of gene expression with WMH using either imputed or measured cell counts. As shown in Additional file 1: Figure S1, the statistics of associations were highly correlated using either the imputed and measured cell counts ( $R^2=0.98$ ), suggesting only marginal effect of imputed cell counts. We also additionally adjusted for the RNA integrity number in our statistical model, and found the results remained largely unchanged (Additional file 1: Table S2). We then performed another sensitive analysis by separating Offspring and Third Generation participants and assessed the association of gene expression with WMH separately. As shown in Additional file 1: Table S3, all 13 top genes had the same direction of association. In addition, the association in Offspring cohort showed relatively stronger effects than that in the Third Generation cohort, reflecting relatively older participants and more WMH volumes comparing to the Third Generation participants.

We also compared the top genes from the current study with previous studies. Among the 13 WMH-related genes, *SEPT11* was also reported to be associated with WMH in brain [36].

### Overlap with GWAS loci

In our recent GWAS that included more than 20,000 participants from diverse ancestries, we identified 4

**Table 2** Most significant genes associated with WMH (FDR < 0.05)

Gene	Primary analysis <sup>a</sup>					Secondary analysis <sup>a</sup>				
	Effect size	SE <sup>b</sup>	P-value	FDR <sup>c</sup>	Direction of effect <sup>d</sup>	Effect size	SE <sup>b</sup>	P-value	Direction of effect <sup>d</sup>	
<i>IL4R</i>	-0.055	0.010	1.5E-08	2.6E-04	↓	-0.050	0.010	3.6E-07	↓	
<i>CD79A</i>	-0.069	0.013	2.7E-07	2.5E-03	↓	-0.067	0.014	1.0E-06	↓	
<i>FCRL6</i>	0.068	0.013	5.2E-07	3.1E-03	↑	0.064	0.014	2.8E-06	↑	
<i>PAX5</i>	-0.052	0.011	3.1E-06	1.4E-02	↓	-0.049	0.011	1.6E-05	↓	
<i>FCRL1</i>	-0.066	0.014	4.8E-06	1.7E-02	↓	-0.059	0.015	5.8E-05	↓	
<i>BANK1</i>	-0.060	0.014	8.6E-06	2.1E-02	↓	-0.057	0.014	3.4E-05	↓	
<i>ARHGAP10</i>	0.029	0.007	9.4E-06	2.1E-02	↑	0.028	0.007	1.9E-05	↑	
<i>YY1</i>	0.018	0.004	9.5E-06	2.1E-02	↑	0.019	0.004	3.5E-06	↑	
<i>TGFBR3</i>	0.047	0.011	1.8E-05	3.6E-02	↑	0.044	0.011	6.4E-05	↑	
<i>IL1RL2</i>	0.021	0.005	2.2E-05	3.7E-02	↑	0.020	0.005	6.1E-05	↑	
<i>SEPT11</i>	0.033	0.008	2.3E-05	3.7E-02	↑	0.034	0.008	1.2E-05	↑	
<i>TREML2</i>	-0.029	0.007	2.6E-05	3.9E-02	↓	-0.026	0.007	2.2E-04	↓	
<i>ARL17A</i>	-0.100	0.024	3.2E-05	4.4E-02	↓	-0.097	0.025	7.8E-05	↓	

<sup>a</sup>Primary analysis was adjusted for age, sex, total cranial volume and cohort, whereas secondary analysis was additional adjusted for smoking, body mass index, systolic blood pressure, diastolic blood pressure, hypertension treatment, total cholesterol, HDL cholesterol, and triglycerides

<sup>b</sup>SE: standard error

<sup>c</sup>FDR: false discovery rate

<sup>d</sup>Direction of effect: ↓ indicates decreased gene expression was associated with increased WMH, whereas ↑ indicates increased gene expression was associated with increased WMH

genetic loci that were significantly associated with WMH [46]. As shown in Table 3, GWAS SNPs at these loci were associated with the expression of 17 genes (FDR < 0.05); three of them was nominally associated with WMH (*SEMA4A*, *UNC13D* and *WBP2*). As an example, the risk allele of rs2984613 was associated with decreased expression of *SEMA4A*, which was associated with increased WMH. We then searched all eQTLs for the 13 WMH-related genes, and found that these eQTLs were significantly enriched with variants associated with WMH [46] ( $P < 2.2e-16$  by Kolmogorov–Smirnov test). Our results suggest that gene expression might serve as an intermediate phenotype that bridges genetic variations and WMH.

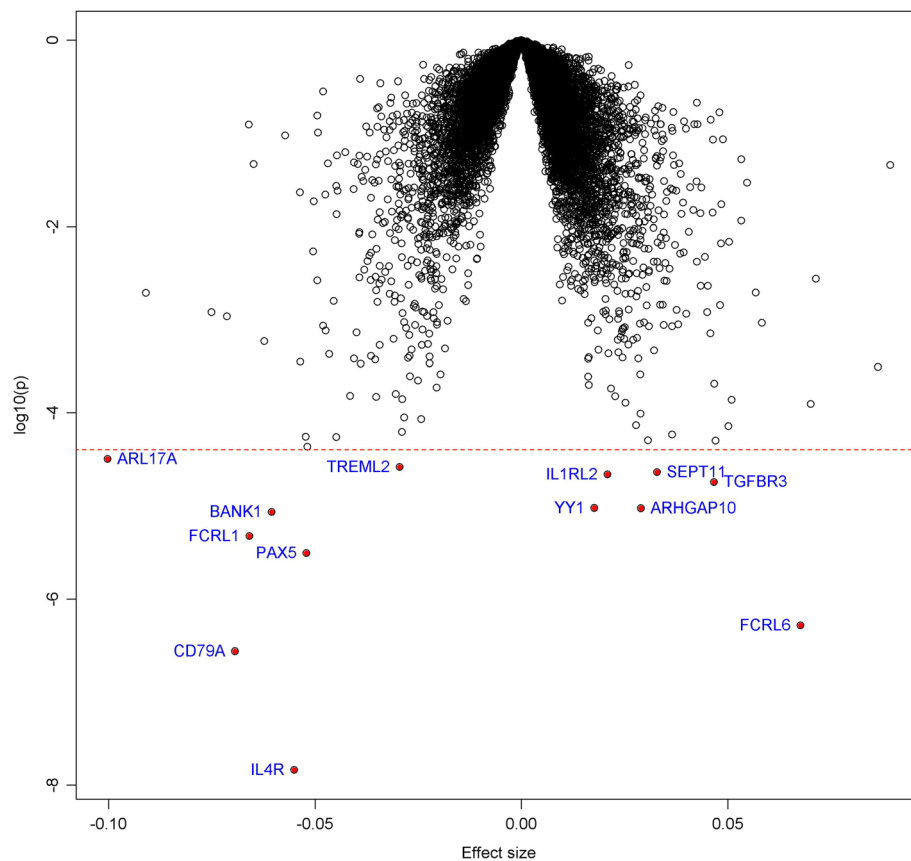
We also examined if previously reported genetic loci for stroke and dementia were associated with 13 WMH-related genes in the current study. GWAS catalog was queried, and 142 genome-wide significant variants ( $P < 5 \times 10^{-8}$ ) were found to be associated with dementia or stroke. We then searched these variants in FHS eQTL database [44] and found that they were associated with the expression of 70 eGenes; 7 of them were nominally associated with WMH, including *ARL17A*, *SYTL2*, *PTGDR*, *POLR2E*, *MS4A6A*, *GPR141*, and *RIN3*. Among them, *ARL17A* was the most significant one and it was associated with SNP rs2732703, which was recently found to be associated with Alzheimer's disease among individuals without APOE  $\epsilon 4$  allele [51].

### Pathway analysis

In order to examine the integrative effects of differentially expressed genes on the biological systems, we examined the enrichment of WMH-related genes in biological pathways using WebGestalt [52]. Given that only 13 genes reached the significance cutoff after correction for multiple testing, we expanded the selection and examined the enrichment of top 1% of genes associated with WMH (including 179 genes). Table 4 shows the top enriched biologic pathways (FDR < 0.05). Many of them are involved in the immune responses and apoptosis, such as antigen processing and presentation (FDR = 0.0019) and apoptosis (FDR = 0.0338).

### Gene interaction network associated with WMH

We applied a dense module searching strategy [48] to construct a WMH-specific subnetwork and examined the interaction between top genes associated with WMH. Note that during the construction of subnetwork, genes with weak or no association with WMH might be also added to the subnetwork if the genes could interact with other significant genes, thus their inclusion would increase the overall score of the subnetwork (see Methods). As shown in Fig. 2, the subnetwork is consisted of 40 nodes and 57 edges, where each node represents one gene, and each edge represents the interaction between two genes. Many of these genes are involved in B cell receptor



**Fig. 1** Volcano plot of gene expression associated with WMH. Each dot represents one gene. The x-axis represents the beta estimation ( $\beta$ ) of each gene, whereas the y-axis represents the  $\log_{10}(P)$ . Positive effects represent that genes were positively associated with WMH, whereas negative effects represent that the genes were negatively associated with WMH. The red dash line indicates  $FDR < 0.05$ . The 13 genes that reached significance cutoff were highlighted

signaling pathway and Epstein-Barr virus infection. *CASP3* appears to be one of the pivotal genes in the network that was connected with 9 other genes, although itself was not associated with WMH ( $P = 0.79$ ). Previously studies also have found that the activation of *CASP3* was observed in brain with ischemic lesions [53–55].

#### Association with neuropsychological performance

The standardized neuropsychological performance testing was described previously [56–58]. Seven matrices were tested, including Hooper Visual Organization Test score, Logical Memories Delayed Recall score, Logical Memories Recognition score, Similarities Test score, Trails Test A score, Trails Test B score, and Trails B–A score. As shown in Additional file 1: Table S4, four genes were significantly associated with at least one cognitive matrix, including *BANK1* for logical memories recognition ( $P = 4.5 \times 10^{-4}$ ), *TGFBR3* for Trails B–A score ( $P = 1.8 \times 10^{-3}$ ) and Trails Test

B score ( $P = 1.9 \times 10^{-3}$ ), *FCRL1* for Hooper Visual Organization Test ( $P = 2.9 \times 10^{-3}$ ) and *FCRL6* for Trails B–A score ( $P = 3.4 \times 10^{-3}$ ). Our results suggest potential shared mechanisms between WMH and cognitive function.

#### Discussion

Increasing evidence has suggested that WMH is an important surrogate of aging and cerebrovascular diseases [26, 28]. However, molecular mechanisms underlying WMH are still poorly understood. In this study, we performed transcriptome profiling on participants who lie across a wide age-range and were ascertained without specific clinical characteristics. We identified 13 genes that were significantly associated with WMH ( $FDR < 0.05$ ) and examined their integrative effect by the network analysis. *CASP3* appears to be one of pivotal genes that interact with multiple genes associated with WMH.



**Table 3** Association of GWAS loci with gene expression. Three genes, *SEMA4A*, *UNC13D*, and *WBP2* were nominally associated with WMH

GWAS SNP	Risk allele <sup>a</sup>	Locus	P value (GWAS association with WMH) [46]	Association of SNPs with gene expression by eQTL analysis [44]			Association of eGene expression with WMH (current study)		
				eGene	P value	Direction on gene expression <sup>b</sup>	Effect size <sup>a</sup>	SE <sup>c</sup>	P value
rs4072479	C	17q25.1	4.6E-14	<i>WBP2</i>	1.0E-10	↓	<b>-0.020</b>	<b>0.010</b>	<b>4.6E-02</b>
rs8067275	T	17q25.1	1.9E-13	<i>MYO15B</i>	2.3E-05	↑	0.003	0.005	0.52
rs9894244	A	17q25.1	5.2E-10	<i>GALK1</i>	2.9E-06	↓	0.001	0.004	0.79
rs9894244	A	17q25.1	5.2E-10	<i>ITGB4</i>	2.9E-06	↓	-0.001	0.003	0.79
rs7216615	C	17q25.1	1.1E-09	<i>H3F3B</i>	9.2E-29	↓	-0.007	0.008	0.38
rs34143128	C	17q25.1	2.4E-09	<i>ACOX1</i>	9.1E-50	↓	0.003	0.009	0.75
rs1135889	A	17q25.1	8.8E-09	<i>UNC13D</i>	4.5E-11	↓	<b>-0.010</b>	<b>0.004</b>	<b>6.1E-03</b>
rs1135889	A	17q25.1	8.8E-09	<i>EXOC7</i>	9.1E-05	↓	-0.009	0.006	0.86
rs1135889	A	17q25.1	8.8E-09	<i>ZACN</i>	9.1E-05	↓	-0.001	0.005	0.86
rs2984613	C	1q22	2.0E-08	<i>C1orf85</i>	6.5E-23	↓	-0.003	0.007	0.65
rs2984613	C	1q22	2.0E-08	<i>TMEM79</i>	6.5E-23	↓	0.005	0.007	0.65
rs2984613	C	1q22	2.0E-08	<i>SMG5</i>	3.3E-20	↓	0.002	0.005	0.72
rs2984613	C	1q22	2.0E-08	<i>PAQR6</i>	1.6E-16	↓	0.001	0.006	0.89
rs2984613	C	1q22	2.0E-08	<i>PMF1</i>	7.3E-10	↑	-0.001	0.005	0.91
rs2984613	C	1q22	2.0E-08	<i>SEMA4A</i>	2.4E-08	↓	<b>-0.022</b>	<b>0.006</b>	<b>3.4E-04</b>
rs10883865	G	10q24.33	4.6E-08	<i>AS3MT</i>	7.9E-26	↓	0.011	0.010	0.30
rs10883865	G	10q24.33	4.6E-08	<i>USMG5</i>	7.7E-07	↑	0.015	0.011	0.19

<sup>a</sup>Risk allele represents the allele that was associated with increased risk of WMH

<sup>b</sup>Direction of effect: ↓ indicates the risk allele was associated with decreased gene expression, whereas ↑ indicates the risk allele was associated with increased gene expression

<sup>c</sup>SE standard error

Genes with P-value less than 0.05 were marked in bold text

Many of the WMH-related genes are involved in the immune response pathway, including the most significant gene, *IL4R* ( $P = 1.5 \times 10^{-8}$ ) and others such as *CD79A*, *TGFBR3* and *IL1RL2*. A variety of studies have suggested a role of inflammatory processes in the development of cerebral large- and small-vessel disease [59, 60]. A high infectious burden is also associated with an increased risk of stroke [61]. Inflammatory markers, such as interleukin-6 and C-reactive protein, have been associated with the presence and progression of white matter lesions across multiple ancestries [24, 62]. Some of top genes are involved in tumorigenesis and Alzheimer type neurodegeneration, which is consistent with prior GWAS.

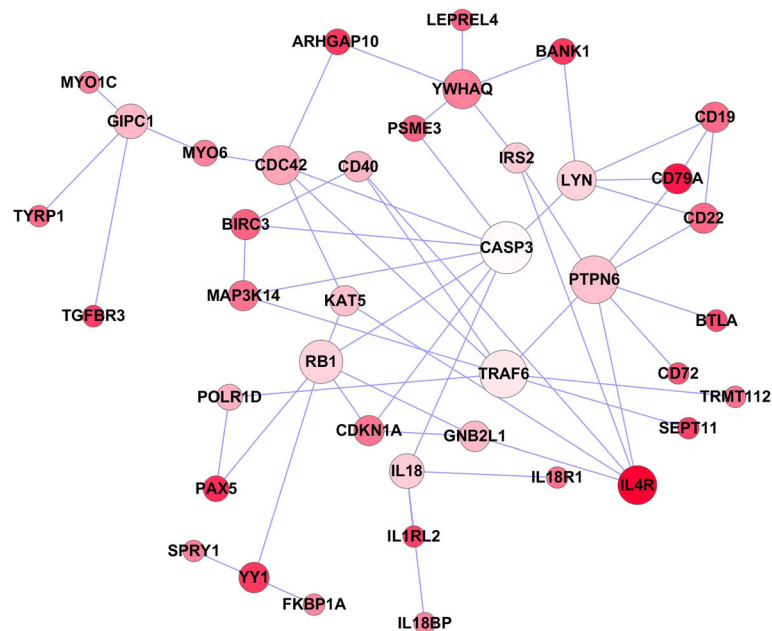
We found a single gene in the current study overlapping with those reported previously [35, 36]. The lack of overlapping might be due to several reasons. The gene expression in the current study was measured by the Affymetrix Exon 1.0 ST array instead of Affymetrix HU133 Plus 2.0 array, which interrogated different sets of genes with different probesets. In addition, our study focused on whole blood samples

but with much larger sample size. Moreover, participants of current study are relatively young and generally healthy, which might represent the WMH burden in the general population. It should be noted that the expression in brain would be more relevant to WMH. However, it is impractical to examine the brain expression in a community-based cohort. We have developed a brain donation program [63], which will be a valuable resource to study brain gene expression profile in future.

We acknowledge several limitations of our study. Gene expression could vary from tissue to tissue over time, but we only measured it in whole blood during a single examination. So we could not study longitudinal changes in gene expression over time, and how this might be related to WMH. Less than half of the studied samples had measured cell counts, and the remaining samples used imputed cell counts, which could introduce some additional variations to our results. In addition, in this cross-sectional study of observational data, we could not infer causal relationships between gene expression and WMH; the observed gene expression changes

**Table 4** Most significant pathways enriched with top 1% genes associated with WMH. Two genes, **IL4R** and **CD79A** were highlighted as they both reached the significance cutoff WMH (FDR < 0.05)

Biological pathway	Total number of genes in the pathway	Ratio of enrichment	P value	FDR	Top 1% genes associated with WMH
Hematopoietic cell lineage	97	7.99	6.0E-06	0.0018	HLA-DOB; HLA-DRA; <b>IL4R</b> ; IL5RA; CD8A; CD19; CD22; CD37
Antigen processing and presentation	77	8.81	1.3E-05	0.0019	PSME3; HLA-DOB; HLA-DRA; KLRC3; KLRD1; CIITA; CD8A
Inflammatory bowel disease (IBD)	65	8.95	5.0E-05	0.0050	HLA-DOB; HLA-DRA; <b>IL4R</b> ; RORA; STAT4; IL18R1
Primary immunodeficiency	37	10.48	5.3E-04	0.0338	CIITA; CD8A; CD19; <b>CD79A</b>
Apoptosis	140	4.85	5.6E-04	0.0338	BCL2L1; TUBA1B; CTSW; GZMB; BIRC3; TNFRSF10C; MAP3K14
Graft-versus-host disease	41	9.46	7.9E-04	0.0399	GZMB; HLA-DOB; HLA-DRA; KLRD1
Cytokine-cytokine receptor interaction	265	2.93	0.005668	0.231	CCR7; <b>IL4R</b> ; IL5RA; PDGFRB; CCL4; CXCR5; TNFRSF10C; IL18R1
B cell receptor signaling pathway	73	5.3	0.006633	0.231	CD19; CD22; CD72; <b>CD79A</b>
Allograft rejection	38	7.7	0.006858	0.231	GZMB; HLA-DOB; HLA-DRA
Type 1 diabetes mellitus	43	6.8	0.009672	0.285	GZMB; HLA-DOB; HLA-DRA



**Fig. 2** WMH-related subnetwork derived from protein-protein interaction. Each node represents one gene, whereas each edge represents the interaction between two genes. The nodes were colored to represent their association with WMH: red color represents strong association, and white color represents no association. The node size is proportional to the number of edges that the node connects to

could be a consequence of vascular injury, both systemic and in the brain. Moreover, all participants included in this study were exclusively adults of European descent. Thus it is unclear if our findings could be generalized to other ethnicities/age groups.

## Conclusions

In conclusion, we performed a large-scale profiling of gene expression in whole blood in a large community-based cohort, and identified 13 genes whose expression was associated with WMH. Our results are consistent with earlier reports that the immune response might be an important pathway to link gene expression and WMH. We also identified genes in glial proliferation and Alzheimer neurodegeneration pathways as potential links to WMH. Future studies with larger sample sizes and better techniques for measurement of gene expression such as RNA sequencing [64, 65] might uncover additional WMH-related genes and novel preventive and therapeutic targets for white matter lesions.

## Additional file

**Additional file 1: Table S1.** Association of top genes with WMH after excluding samples with stroke, dementia or vascular diseases. **Table S2.** Association of top genes with WMH after additionally adjusted for the RNA integrity number (RIN). **Table S3.** Separated analysis for participants from the Offspring cohort and the Third Generation cohort. **Table S4.** Association of top genes within cognitive performance. **Figure S1.** Correlation between the statistics of WMH associations derived from the

imputed cell counts or the measured cell counts using those samples who have measured cell counts. X-axis represents the statistics derived from measured cell counts, while y-axis represents the statistics derived from the imputed cell counts. Strong correlation was observed ( $R^2 = 0.984$ ), suggesting only marginal effect of imputed cell counts. (PDF 187 kb)

## Abbreviations

eQTL: Expression quantitative trait loci; FDR: False discovery rate; FHS: Framingham Heart Study; GWAS: Genome-wide association study; TCV: Total intracranial volume; WMH: White matter hyperintensities

## Acknowledgements

Not applicable

## Funding

FHS gene expression profiling was funded through the Division of Intramural Research (Principal Investigator, Daniel Levy), National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD. The Framingham Heart Study is supported by National Heart, Lung, and Blood Institute contract HHSN2682015000011. Additional support was provided by grants from the National Institute on Aging (R01 AG033193, R01 AG008122, U01 AG049505, R01 AG049607, R01 AG054076) and the National Institute on Neurological Disorders and Stroke (R01 NS017950 and UH2 NS100605). None of the funding agencies had any role in shaping the analyses or presentation of results.

## Availability of data and materials

The gene expression data in the study is available at dbGap [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000363.v16.p10](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000363.v16.p10).

## Authors' contributions

HL and SS initiated the study and drafted the manuscript. HL, CS, ZX, QY, CW, AB and SS analyzed and interpreted the result. TH, RJ, PJM and DL generated and processed the gene expression data. All authors critically reviewed and approved the final manuscript.



**Ethics approval and consent to participate**

All participants gave written informed consent, and the study was approved by the institutional review board at the Boston University Medical Center.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>National Heart Lung and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA, USA. <sup>2</sup>Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, 72 East Concord Street, B-616, Boston, MA 02118, USA. <sup>3</sup>Department of Neurology, Boston University School of Medicine, 72 East Concord Street, B-602, Boston, MA 02118, USA. <sup>4</sup>College of Basic Medical Science, Zhejiang Chinese Medical University, Hangzhou, Zhejiang Province, China. <sup>5</sup>Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. <sup>6</sup>Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD, USA. <sup>7</sup>Mathematical and Statistical Computing Laboratory, Center for Information Technology, National Institute of Health, Bethesda, MD, USA. <sup>8</sup>Hebrew Senior Life, 1200 Centre Street Room #609, Boston, MA 02131, USA.

Received: 30 May 2017 Accepted: 12 September 2017

Published online: 18 September 2017

**References**

- DeBette S, Markus HS. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ*. 2010;341:c3666.
- de Leeuw FE, de Groot JC, Bots ML, Witteman JC, Oudkerk M, Hofman A, van Gijn J, Breteler MM. Carotid atherosclerosis and cerebral white matter lesions in a population based magnetic resonance imaging study. *Journal of neurology*. 2000;247:291–6.
- de Leeuw FE, de Groot JC, Achten E, Oudkerk M, Ramos LM, Heijboer R, Hofman A, Jolles J, van Gijn J, Breteler MM. Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study. The Rotterdam scan study. *J Neurol Neurosurg Psychiatry*. 2001;70:9–14.
- Ylikoski A, Erkinjuntti T, Raininko R, Sarna S, Sulkava R, Tilvis R. White matter hyperintensities on MRI in the neurologically nondiseased elderly. Analysis of cohorts of consecutive subjects aged 55 to 85 years living at home. *Stroke*. 1995;26:1171–7.
- Garde E, Mortensen EL, Krabbe K, Rostrup E, Larsson HB. Relation between age-related decline in intelligence and cerebral white-matter hyperintensities in healthy octogenarians: a longitudinal study. *Lancet*. 2000;356:628–34.
- Longstreth WT Jr. Brain vascular disease overt and covert. *Stroke*. 2005;36:2062–3.
- Launer LJ. Epidemiology of white matter lesions. *Top Magn Reson Imaging*. 2004;15:365–7.
- Ylikoski R, Ylikoski A, Erkinjuntti T, Sulkava R, Raininko R, Tilvis R. White matter changes in healthy elderly persons correlate with attention and speed of mental processing. *Archives Neurol*. 1993;50:818–24.
- Au R, Massaro JM, Wolf PA, Young ME, Beiser A, Seshadri S, D'Agostino RB, DeCarli C. Association of white matter hyperintensity volume with decreased cognitive functioning: the Framingham heart study. *Archives of neurology*. 2006;63:246–50.
- DeBette S, Beiser A, DeCarli C, Au R, Himali JJ, Kelly-Hayes M, Romero JR, Kase CS, Wolf PA, Seshadri S. Association of MRI markers of vascular brain injury with incident stroke, mild cognitive impairment, dementia, and mortality: the Framingham offspring study. *Stroke*. 2010;41:600–6.
- Kuller LH, Lopez OL, Newman A, Beauchamp NJ, Burke G, Dulberg C, Fitzpatrick A, Fried L, Haan MN. Risk factors for dementia in the cardiovascular health cognition study. *Neuroepidemiology*. 2003;22:13–22.
- Meguro K, Ishii H, Kasuya M, Akanuma K, Meguro M, Kasai M, Lee E, Hashimoto R, Yamaguchi S, Asada T. Incidence of dementia and associated risk factors in Japan: the Osaki-Tajiri project. *J Neurol Sci*. 2007;260:175–82.
- Bigler ED, Lowry CM, Kerr B, Tate DF, Hessel CD, Earl HD, Miller MJ, Rice SA, Smith KH, Tschanz JT, et al. Role of white matter lesions, cerebral atrophy, and APOE on cognition in older persons with and without dementia: the Cache County, Utah, study of memory and aging. *Neuropsychology*. 2003; 17:339–52.
- Burton EJ, Kenny RA, O'Brien J, Stephens S, Bradbury M, Rowan E, Kalaria R, Firbank M, Wesnes K, Ballard C. White matter hyperintensities are associated with impairment of memory, attention, and global cognitive performance in older stroke patients. *Stroke*. 2004;35:1270–5.
- Gunning-Dixon FM, Raz N. Neuroanatomical correlates of selected executive functions in middle-aged and older adults: a prospective MRI study. *Neuropsychologia*. 2003;41:1929–41.
- Gerdes VE, Kwa VI, ten Cate H, Brandjes DP, Buller HR, Stam J. Amsterdam vascular medicine G: cerebral white matter lesions predict both ischemic strokes and myocardial infarctions in patients with established atherosclerotic disease. *Atherosclerosis*. 2006;186:166–72.
- Bokura H, Kobayashi S, Yamaguchi S, Iijima K, Nagai A, Toyoda G, Oguro H, Takahashi K. Silent brain infarction and subcortical white matter lesions increase the risk of stroke and mortality: a prospective cohort study. *J Stroke Cerebrovasc Dis*. 2006;15:57–63.
- Wong TY, Klein R, Sharrett AR, Couper DJ, Klein BE, Liao DP, Hubbard LD, Mosley TH. Study AIARIC: cerebral white matter lesions, retinopathy, and incident clinical stroke. *JAMA*. 2002;288:67–74.
- Buyck JF, Dufouil C, Mazoyer B, Maillard P, Ducimetiere P, Alperovitch A, Bousser MG, Kurth T, Tzourio C. Cerebral white matter lesions are associated with the risk of stroke but not with other vascular events: the 3-City Dijon study. *Stroke*. 2009;40:2327–31.
- Kuller LH, Arnold AM, Longstreth WT Jr, Manolio TA, O'Leary DH, Burke GL, Fried LP, Newman AB. White matter grade and ventricular volume on brain MRI as markers of longevity in the cardiovascular health study. *Neurobiol Aging*. 2007;28:1307–15.
- Ikram MA, Vernooij MW, Vrooman HA, Hofman A, Breteler MM. Brain tissue volumes and small vessel disease in relation to the risk of mortality. *Neurobiol Aging*. 2009;30:450–6.
- Yamauchi H, Fukuda H, Oyanagi C. Significance of white matter high intensity lesions as a predictor of stroke from arteriosclerosis. *J Neurol Neurosurg Psychiatry*. 2002;72:576–82.
- Schmidt R, Scheltens P, Erkinjuntti T, Pantoni L, Markus HS, Wallin A, Barkhof F, Fazekas F. White matter lesion progression: a surrogate endpoint for trials in cerebral small-vessel disease. *Neurology*. 2004;63:139–44.
- Fornage M, Chiang YA, O'Meara ES, Psaty BM, Reiner AP, Siscovick DS, Tracy RP, Longstreth WT Jr. Biomarkers of inflammation and MRI-defined small vessel disease of the brain: the cardiovascular health study. *Stroke*. 2008;39:1952–9.
- Carmelli D, DeCarli C, Swan GE, Jack LM, Reed T, Wolf PA, Miller BL. Evidence for genetic variance in white matter hyperintensity volume in normal elderly male twins. *Stroke*. 1998;29:1177–81.
- Atwood LD, Wolf PA, Heard-Costa NL, Massaro JM, Beiser A, D'Agostino RB, DeCarli C. Genetic variation in white matter hyperintensity volume in the Framingham study. *Stroke; a journal of cerebral circulation*. 2004;35:1609–13.
- Turner ST, Jack CR, Fornage M, Mosley TH, Boerwinkle E, de Andrade M. Heritability of leukoaraiosis in hypertensive sibships. *Hypertension*. 2004;43:483–7.
- DeStefano AL, Atwood LD, Massaro JM, Heard-Costa N, Beiser A, Au R, Wolf PA, DeCarli C. Genome-wide scan for white matter hyperintensity: the Framingham heart study. *Stroke*. 2006;37:77–81.
- Paternoster L, Chen W, Sudlow CL. Genetic determinants of white matter hyperintensities on brain scans: a systematic assessment of 19 candidate gene polymorphisms in 46 studies in 19,000 subjects. *Stroke*. 2009;40:2020–6.
- Turner ST, Fornage M, Jack CR Jr, Mosley TH, Kardia SL, Boerwinkle E, de Andrade M. Genomic susceptibility loci for brain atrophy in hypertensive sibships from the GENOA study. *Hypertension*. 2005;45:793–8.
- Fornage M, DeBette S, Bis JC, Schmidt H, Ikram MA, Dufouil C, Sigurdsson S, Lumley T, DeStefano AL, Fazekas F, et al. Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. *Annals Neurology*. 2011;69:928–39.
- Lin H, Yin X, Lunetta KL, Dupuis J, McManus DD, Lubitz SA, Magnani JW, Joehanes R, Munson PJ, Larson MG, et al. Whole blood gene expression and atrial fibrillation: the Framingham heart study. *PLoS ONE*. 2014;9:e96794.

33. Lin H, Joehanes R, Pilling LC, Dupuis J, Lunetta KL, Ying SX, Benjamin EJ, Hernandez D, Singleton A, Melzer D, et al. Whole blood gene expression and interleukin-6 levels. *Genomics*. 2014;104:490–5.
34. Joehanes R, Ying S, Huan T, Johnson AD, Raghavachari N, Wang R, Liu P, Woodhouse KA, Sen SK, Tanriverdi K, et al. Gene expression signatures of coronary heart disease. *Arterioscler Thromb Vasc Biol*. 2013;33:1418–26.
35. Xu H, Stamova B, Jickling G, Tian Y, Zhan X, Ander BP, Liu D, Turner R, Rosand J, Goldstein LB, et al. Distinctive RNA expression profiles in blood associated with white matter hyperintensities in brain. *Stroke; a journal of cerebral circulation*. 2010;41:2744–9.
36. Simpson JE, Hosny O, Wharton SB, Heath PR, Holden H, Fernando MS, Matthews F, Forster G, O'Brien JT, Barber R, et al. Microarray RNA expression analysis of cerebral white matter lesions reveals changes in multiple functional pathways. *Stroke*. 2009;40:369–75.
37. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham study. *Ann N Y Acad Sci*. 1963;107:539–56.
38. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Preventive Med*. 1975;4:518–25.
39. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979;110:281–90.
40. DeCarli C, Massaro J, Harvey D, Hald J, Tullberg M, Au R, Beiser A, D'Agostino R, Wolf PA. Measures of brain morphology and infarction in the framingham heart study: establishing what is normal. *Neurobiology of aging*. 2005;26:491–510.
41. DeCarli C, Fletcher E, Ramey V, Harvey D, Jagust WJ. Anatomical mapping of white matter hyperintensities (WMH): exploring the relationships between periventricular WMH, deep WMH, and total WMH burden. *Stroke*. 2005;36:50–5.
42. Carmichael O, Mungas D, Beckett L, Harvey D, Tomaszewski Farias S, Reed B, Olichney J, Miller J, Decarli C. MRI predictors of cognitive change in a diverse and carefully characterized elderly population. *Neurobiol Aging*. 2012;33:83–95.
43. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4:249–64.
44. Joehanes R, Zhang X, Huan T, Yao C, Ying SX, Nguyen QT, Demirkale CY, Feolo ML, Sharopova NR, Sturcke A, et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biology*. 2017;18:16.
45. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc*. 1995;57:289–300.
46. Verhaaren BF, DeBette S, Bis JC, Smith JA, Ikram MK, Adams HH, Beecham AH, Rajan KB, Lopez LM, Barral S, et al. Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circ Cardiovasc Genet*. 2015;8:398–409.
47. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res*. 2017;45:D896–901.
48. Jia P, Zheng S, Long J, Zheng W, Zhao Z: dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics*. 2011;27:95–102.
49. Cowley MJ, Pinese M, Kassahn KS, Waddell N, Pearson JV, Grimmond SM, Biankin AV, Hautaniemi S, Wu J. PINA v2.0: mining interactome modules. *Nucleic Acids Res*. 2012;40:D862–5.
50. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*. 2002; 18(Suppl 1):S233–40.
51. Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert JC, Chung J, Naj AC, Kunkle BW, Wang LS, Bis JC, Bellenguez C, et al. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol Psychiatry*. 2016;21:108–17.
52. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GENE SeT Analysis toolkit (WebGestalt): update 2013. *Nucleic Acids Res*. 2013;41:W77–83.
53. Gorup D, Bohacek I, Milicevic T, Pochet R, Mitrecic D, Kriz J, Gajovic S. Increased expression and colocalization of GAP43 and CASP3 after brain ischemic lesion in mouse. *Neuroscience Lett*. 2015;597:176–82.
54. Manabat C, Han BH, Wendland M, Derugin N, Fox CK, Choi J, Holtzman DM, Ferriero DM, Vexler ZS. Reperfusion differentially induces caspase-3 activation in ischemic core and penumbra after stroke in immature brain. *Stroke*. 2003;34:207–13.
55. Didenko W, Ngo H, Minchew CL, Boudreaux DJ, Widmayer MA, Baskin DS. Caspase-3-dependent and -independent apoptosis in focal brain ischemia. *Mol Med*. 2002;8:347–52.
56. Au R, Seshadri S, Wolf PA, Elias M, Elias P, Sullivan L, Beiser A, D'Agostino RB. New norms for a new generation: cognitive performance in the framingham offspring cohort. *Exp Aging Res*. 2004;30:333–58.
57. Therkelsen KE, Preis SR, Beiser A, DeCarli C, Seshadri S, Wolf P, Au R, Fox CS. Neck circumference, brain imaging measures, and neuropsychological testing measures. *J Stroke Cerebrovasc Dis*. 2016;25:1570–81.
58. Nishtala A, Preis SR, Beiser A, Devine S, Hanke L, Seshadri S, Wolf PA, Au R. Midlife cardiovascular risk impacts executive function: Framingham offspring study. *Alzheimer Dis Assoc Disord*. 2014;28:16–22.
59. Simpson JE, Ince PG, Higham CE, Gelsthorpe CH, Fernando MS, Matthews F, Forster G, O'Brien JT, Barber R, Kalaria RN, et al. Microglial activation in white matter lesions and nonlesional white matter of ageing brains. *Neuropathol Appl Neurobiol*. 2007;33:670–83.
60. Fassbender K, Bertsch T, Mielke O, Muhlhauser F, Hennerici M. Adhesion molecules in cerebrovascular diseases. Evidence for an inflammatory endothelial activation in cerebral large- and small-vessel disease. *Stroke*. 1999;30:1647–50.
61. Palm F, Pussinen PJ, Aigner A, Becher H, Bugge F, Bauer MF, Grund-Ginsbach C, Safer A, Urbanek C, Grau AJ. Association between infectious burden, socioeconomic status, and ischemic stroke. *Atherosclerosis*. 2016;254:117–23.
62. Nagai K, Kozaki K, Sonohara K, Akishita M, Toba K. Relationship between interleukin-6 and cerebral deep white matter and periventricular hyperintensity in elderly women. *Geriatr Gerontol Int*. 2011;11:328–32.
63. Au R, Seshadri S, Knox K, Beiser A, Himali JJ, Cabral HJ, Auerbach S, Green RC, Wolf PA, McKee AC. The Framingham brain donation program: neuropathology along the cognitive continuum. *Curr Alzheimer Res*. 2012;9:673–86.
64. Adam O, Lavall D, Theobald K, Hohl M, Grube M, Ameling S, Sussman MA, Rosenkranz S, Kroemer HK, Schafers HJ, et al. Rac1-induced connective tissue growth factor regulates connexin 43 and N-cadherin expression in atrial fibrillation. *J Am Coll Cardiol*. 2010;55:469–80.
65. Matkovich SJ, Zhang Y, Van Booven DJ, Dorn GW 2nd. Deep mRNA sequencing for in vivo functional analysis of cardiac transcriptional regulators: application to Galphaq. *Circ Res*. 2010;106:1459–67.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
www.biomedcentral.com/submit

