Translational Stroke Research (2019) 10:26–35
https://doi.org/10.1007/s12975-018-0623-1

ORIGINAL ARTICLE

Shifts in Leukocyte Counts Drive the Differential Expression of Transcriptional Stroke Biomarkers in Whole Blood

Grant C. O'Connell[1] · Madison B. Treadway[2] · Connie S. Tennant[3] · Noelle Lucke-Wold[3] · Paul D. Chantler[4,5] · Taura L. Barr[6]

Received: 11 November 2017 / Revised: 12 February 2018 / Accepted: 9 March 2018 / Published online: 17 March 2018

Abstract

Our group recently identified a panel of ten genes whose RNA expression levels in whole blood have utility for detection of stroke. The purpose of this study was to determine the mechanisms by which these genes become differentially expressed during stroke pathology. First, we assessed the transcriptional distribution of the ten genes across the peripheral immune system by measuring their expression levels on isolated neutrophils, monocytes, B-lymphocytes, CD-4+ T-lymphocytes, CD-8+ T-lymphocytes, and NK-cells generated from the blood of healthy donors ($n = 3$). Then, we examined the relationship between the whole-blood expression levels of the ten genes and white blood cell counts in a cohort of acute ischemic stroke patients ($n = 36$) and acute stroke mimics ($n = 15$) recruited at emergency department admission. All ten genes displayed strong patterns of lineage-specific expression in our analysis of isolated leukocytes, and their whole-blood expression levels were correlated with white blood cell differential across the total patient population, suggesting that many of them are likely differentially expressed in whole blood during stroke as an artifact of stroke-induced shifts in leukocyte counts. Specifically, factor analysis inferred that over 50% of the collective variance in their whole-blood expression levels across the patient population was driven by underlying variance in white blood cell counts alone. However, the cumulative expression levels of the ten genes displayed a superior ability to discriminate between stroke patients and stroke mimics relative to white blood cell differential, suggesting that additional less prominent factors influence their expression levels which add to their diagnostic utility. These findings not only provide insight regarding this particular panel of ten genes, but also into the results of prior stroke transcriptomics studies performed in whole blood.

Keywords Triage · Microarray · RNA sequencing · CBC · Complete blood count · PCR · NLR · Neutrophil-lymphocyte ratio · MLR · Monocyte-lymphocyte ratio

Introduction

The identification of circulating biomarkers which could be used to accurately detect stroke would allow for the development of molecular diagnostics which could better inform triage and treatment decisions in the acute phase of care. The peripheral immune system responds robustly to stroke, and multiple genome-wide transcriptomic studies have

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12975-018-0623-1) contains supplementary material, which is available to authorized users.

✉ Grant C. O'Connell
grant.oconnell@case.edu

[1] School of Nursing, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4904, USA

[2] Department of Biology, Eberly College of Arts and Sciences, West Virginia University, Morgantown, WV, USA

[3] Center for Basic and Translational Stroke Research, Robert C. Byrd Health Sciences Center, West Virginia University, Morgantown, WV, USA

[4] Center for Cardiovascular and Respiratory Sciences, Robert C. Byrd Health Sciences Center, West Virginia University, Morgantown, WV, USA

[5] Division of Exercise Physiology, School of Medicine, West Virginia University, Morgantown, WV, USA

[6] Valtari Bio Incorporated, Morgantown, WV, USA

demonstrated that stroke induces widespread alterations in gene expression at the level of whole blood [1–5]. Recently, our group leveraged this phenomenon to identify a blood-based RNA panel consisting of seven upregulated genes (*PLXDC2*, *STK3*, *ANTXR2*, *KIF1B*, *CD163*, *PDK4*, *CTSZ*) and three downregulated genes (*GRAP*, *MAL*, *ID3*) which has demonstrated the ability to accurately detect acute ischemic stroke (AIS) in three independent patient populations [3, 6]. While this candidate biomarker panel has potential for clinical use, the mechanisms by which its constituent genes become differentially expressed in whole blood during stroke remain unclear.

Stroke induces rapid and pronounced changes in the complexion of the circulating leukocyte pool. Increases in damage-associated molecular signaling at the site of injury, along with elevations in sympathetic innervation of peripheral myeloid tissue, trigger robust activation of the innate immune system [7–9], resulting in a substantial rise in circulating monocyte and neutrophil counts in the hours following onset [10, 11]. Simultaneously, sympathetic signaling to peripheral lymphoid organs shifts the adaptive immune system into a state of suppression [7–9], which results in a marked reduction in circulating lymphocyte counts that can persist for weeks post-injury [11]. Within the context of this immune response, it is likely that these genes exhibit differential expression in whole blood during stroke as a result of two not-necessarily distinct scenarios:

First, it is possible that these genes undergo true transcriptional modifications at the level of individual circulating cells, and that these transcriptional changes help drive the stroke-induced shift in peripheral immune phenotype. This scenario is intriguing from the standpoint that the stroke-induced peripheral immune response significantly influences patient outcome, as alterations in its magnitude and duration are known to drive secondary complications such as hemorrhagic transformation and post-stroke infection [10–15]; if these genes are biological drivers of stroke immunopathology, it would make them legitimate therapeutic targets which could be leveraged for immunomodulation.

Conversely, it is possible that these genes exhibit altered whole-blood expression in response to stroke simply as artifacts of the stroke-induced change in leukocyte counts. The numerous subpopulations of leukocytes which comprise the peripheral immune system carry transcriptomes which are highly unique [16]. Thus, shifts in white blood cell (WBC) differential can result in altered gene expression at the level of whole blood, even in the absence of true transcriptional modifications at the level of individual cells [17]; this is especially true in the case of genes which exhibit extreme transcriptional heterogeneity across leukocyte subpopulations. If the genes in question are differentially regulated in whole blood as a result of this scenario, they would still have diagnostic utility; however, it would be less likely that they are biological drivers of pathology.

Not only could identifying which of the aforementioned mechanisms drive the differential expression of these genes improve their clinical utility as biomarkers by providing biological context, but it could also produce novel insight into stroke immunopathology as a whole. Thus, the purpose of this study was to examine the potential mechanisms by which this previously identified ten-marker panel of stroke-associated genes becomes differentially expressed in whole blood during stroke.

Materials and Methods

Experimental Design

Peripheral venous blood samples were obtained from 36 AIS patients and 15 acute stroke mimics at emergency department admission (Table 1). In addition, liquid biopsies were also obtained from three healthy donors. In order to examine the susceptibility of the ten previously identified stroke-associated genes to undergo altered expression in whole blood as a result of perturbations in WBC differential, we first examined their transcriptional distribution across numerous isolated subpopulations of leukocytes generated from healthy donor peripheral blood. We then assessed the individual relationships between the whole-blood RNA expression levels of each of the ten genes and WBC differential across the total patient population; factor analysis was subsequently used to estimate what proportion of variance in their collective expression levels was attributable to WBC differential. Lastly, we compared the ability of whole-blood expression levels of the ten genes to discriminate between stroke patients and stroke mimics to that of WBC differential alone.

Subjects

Healthy donors, AIS patents, and stroke mimics were recruited from 2011 to 2015 at Ruby Memorial Hospital, Morgantown, WV. All AIS patients displayed definitive radiographic evidence of vascular ischemic pathology on magnetic resonance imaging (MRI) or computed tomography (CT) according to the criteria for diagnosis of acute ischemic cerebrovascular syndrome (AICS) [18]. Patients admitted to the emergency department as suspected strokes based on the overt presentation of stroke-like symptoms but receiving a definitive negative AICS diagnosis upon neuroradiological imaging and clinical evaluation were identified as acute stroke mimics [18]. Discharge diagnoses of stroke mimics included cases of seizures, complex migraines, hypertensive encephalopathy, and other non-stroke conditions which induce neurological symptoms. All diagnoses were confirmed by an experienced neurologist. Patients were excluded from enrolment if they received a non-definitive AICS diagnosis, presented with evidence of hemorrhagic pathology on neuroradiological

Table 1 Clinical and demographic characteristics

	Mimic ($n = 15$)	Stroke ($n = 36$)	p
[a]Age (mean ± SD)	60.2 ± 17.2	73.1 ± 12.85	0.005*
[b]Female n (%)	8 (53.3)	22 (61.1)	0.757
[a]NIHSS (mean ± SD)	5.0 ± 4.5	9.0 ± 7.6	0.064
Partial anterior cerebral infarction n (%)	–	19 (52.8)	–
Total anterior cerebral infarction n (%)	–	5 (13.9)	–
Posterior cerebral infarction n (%)	–	8 (22.2)	–
Lacunar cerebral infarction n (%)	–	4 (11.1)	–
[c]Infarct volume (mean ± SD)	–	29.3 ± 47.2	–
[a,d]Time from onset to blood draw (mean ± SD)	368 ± 325	438 ± 326	0.492
[b]Family history of stroke n (%)	4 (26.7)	14 (38.9)	0.527
[b]Hypertension n (%)	13 (86.7)	30 (83.3)	1.000
[b]Dyslipidemia n (%)	10 (66.7)	16 (44.4)	0.220
[b]Diabetes n (%)	5 (33.3)	7 (19.4)	0.302
[b]Previous stroke n (%)	4 (26.7)	7 (19.4)	0.711
[b]Atrial fibrillation n (%)	3 (20.0)	12 (33.3)	0.503
[b]Myocardial infarction n (%)	5 (33.3)	10 (27.7)	0.743
[b]Hypertension medication n (%)	13 (86.7)	25 (69.4)	0.297
[b]Diabetes medication n (%)	5 (33.3)	7 (19.4)	0.302
[b]Cholesterol medication n (%)	9 (60.0)	13 (36.1)	0.135
[b]Anticoagulant or antiplatelet n (%)	11 (73.3)	22 (61.1)	0.527
[b]Current smoking n (%)	1 (6.70)	8 (22.2)	0.251

[a] Statistical comparison via two-sample two-tailed t test

[b] Statistical comparison via 2 × 2 Fisher's exact test

[c] Measured in cubic centimeters

[d] Measured in minutes

imaging, presented with cancer as a co-pathology, were under 18 years of age, or arrived to the emergency department outside of 24 h from symptom onset. Injury severity was determined according to the National Institutes of Health stroke scale (NIHSS). Demographic information was collected from either the subject or significant other by a trained clinician. Procedures were approved by the institutional review boards of West Virginia University and Ruby Memorial Hospital (IRB protocol 1410450461R001), and written informed consent was obtained from all subjects or their authorized representatives prior to study procedures.

Neuroradiological Imaging

CT imaging was performed on either a Toshiba Aquilion 64 or Toshiba Aquilion 320 scanner (5 mm slices acquired at 120 KVp, 250 mA). MRI imaging was performed using either a Siemens Aera or Siemens Verio scanner (5 mm slices, 5 mm inter-slice gap, acquired via diffusion-weighted echoplanar imaging with a B-factor of 1000 s/mm^2). The BrainLAB iPlan Neuroradiology software package (BrainLAB, Westchester, Ill) was used to calculate infarct volume via manual tracing.

Type of cerebral infarction was classified via the four categories described by Bamford et al. based solely on radiographic evidence of the following: anterior circulation infarct with restricted cortical involvement (partial anterior cerebral infarction), large anterior circulation infarct with both cortical and subcortical involvement (total anterior cerebral infarction), small perforating artery infarction (lacunar cerebral infarction), and vertebrobasilar territory infarction (posterior cerebral infarction) [19, 20].

Blood Collection

All blood sampling was performed prior to the administration of thrombolytics. Parallel peripheral venous whole-blood samples were collected from subjects via PAXgene RNA tubes (Qiagen, Valencia, CA) and K_2EDTA Vacutainers (Becton Dickenson, Franklin Lakes, NJ). PAXgene RNA tubes were frozen immediately and stored at − 80 °C until RNA extraction, while K_2EDTA tubes were stored at room temperature until white blood cell differential (less than 30 min) or leukocyte isolation (performed immediately).

Isolation of Leukocyte Subpopulations

Leukocytes were isolated from EDTA-treated blood via immuno-magnetic negative selection (EasySep Direct, StemCell Technologies, Vancouver, BC). Neutrophils, monocytes, CD4+ T-lymphocytes, CD8+ T-lymphocytes, B-lymphocytes, and NK-cells were isolated from 4 mL of blood per population according to the manufacture-recommended protocol. Isolated cells were rinsed once in PBS and lysed in Qiagen buffer RLT containing beta-mercaptoethanol, flash frozen in liquid nitrogen, and stored at − 80 °C until RNA extraction.

RNA Extraction and qRT-PCR

Whole-blood RNA was extracted from PAXgene tubes via the PreAnalytiX PAXgene blood RNA kit (Qiagen) and automated using the QIAcube system (Qiagen). RNA was isolated from leukocyte lysates via the RNeasy Micro kit (Qiagen). Quantity and purity of isolated RNA was determined via spectrophotometry (NanoDrop, Thermo Scientific, Waltham, MA). cDNA was generated from purified RNA via Applied Biosystems high-capacity reverse transcription kit. For qPCR, target sequences were amplified from 10 ng of cDNA input using sequence-specific primers (Supplemental Table 1) and detected via SYBR green (PowerSYBR, Thermo-Fisher) on the RotorGeneQ (Qiagen). Raw amplification plots were background corrected and CT values were generated via the RotorGeneQ software package. All reactions were performed in triplicate. Transcripts of *B2M*, *PPIB*, and *ACTB* were amplified as low-variability references, and normalization was performed using the NORMA-gene data-driven normalization algorithm [21, 22].

White Blood Cell Differential and NMLR

Complete blood count was obtained from EDTA-treated blood via combined optical flow cytometry and cellular impedance using an automated clinical hematology analyzer (Cell-Dyn, Abbott Diagnostics, Santa Clara, CA). WBC differential was summarized in terms of the neutrophil-monocyte to lymphocyte ratio (NMLR), as it provides a single variable which can be intuitively used to infer peripheral immune status within the context of the known shifts in WBC counts which occur in response to stroke. NMLR was calculated as the geometric mean of absolute neutrophil and absolute monocyte count divided by absolute lymphocyte count:

$$\mathrm{NMLR} = \frac{\sqrt{\text{neutrophil count}*\text{monocyte count}}}{\text{lymphocyte count}}$$

Statistics

All statistics were performed using R 3.3 (R project for statistical computing) [23]. Fisher's exact test was used for comparison of dichotomous variables, while *t* test or one-way ANOVA was used for the comparison of continuous variables where appropriate. Either Spearman's rho or Pearson's *r* was used to assess the strength of correlational relationships depending on level of homoscedasticity. Factor analysis was performed using the factanal() function; prior to analysis, data were tested for sphericity and sampling adequacy via Bartlett's test and the Kaiser-Mayer-Olkin index. Covariance matrix hierarchical clustering was performed using the hclust() function.

The diagnostic ability of binary classifiers was assessed via receiver operator characteristic (ROC) using the "pROC" package [24]. In the case of classification using a single variable, ROC curves were generated directly using the roc() function. In the case of classification using multiple variables, classification was performed using k nearest neighbors (kNN) via the knn.cv() function of the "class" package [25], and the resultant prediction probabilities were used for ROC curve generation. Statistical comparison of area under ROC curves was performed using roc.test() function according the non-parametric method described by DeLong et al. [26].

Sample sizes were arbitrarily determined. In the case of all statistical testing, the null hypothesis was rejected when $p < 0.05$. In the case of multiple comparisons, *P* values were adjusted via the Benjamini-Hochberg method [27]. The parameters of all statistical tests performed are outlined in detail within the figure legends.

Results

Stroke-Associated Genes Exhibit Lineage-Specific Expression in the Peripheral Immune System

In order to examine the susceptibility of the ten previously identified stroke-associated genes to undergo altered expression in whole blood as a result of perturbations in WBC differential, we first examined their transcriptional distribution across numerous isolated subpopulations of leukocytes generated from the peripheral blood of healthy donors. Interestingly, each of the ten genes exhibited some degree of lineage-specific expression. All seven genes which have been previously identified as being upregulated in whole blood during AIS were predominantly expressed on cell populations of myeloid origin. Conversely, all three genes which have been previously identified as being downregulated in whole blood during AIS were predominantly expressed on cells of lymphoid origin (Fig. 1). In the case of several genes, the differences in expression between myeloid and lymphoid populations were quite substantial, ranging from hundreds-fold to thousands-fold; this extreme heterogeneity in expression suggests that the whole blood expression levels of these genes could be dramatically influenced by shifts in white blood cell differential.

Whole-Blood Expression Levels of Stroke-Associated Genes Are Strongly Correlated with WBC Differential

Next, we examined the relationship between the whole-blood expression levels of the ten previously identified stroke-associated genes and WBC differential in liquid biopsies obtained from AIS patients and stroke mimics at emergency

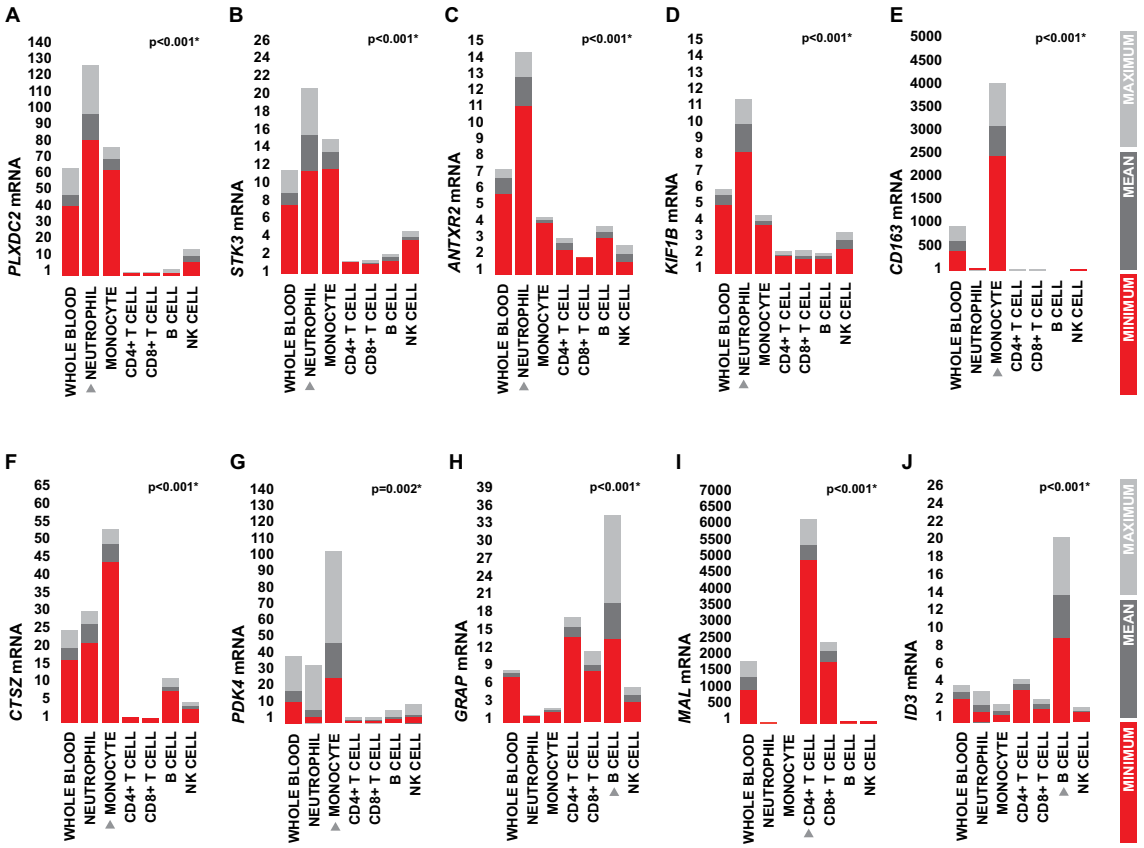

Fig. 1 Transcriptional distribution of stroke-associated genes across the peripheral immune system. Expression levels of the ten stroke-associated genes on isolated leukocyte subpopulations generated from the peripheral blood of one male and two female healthy donors. Expression levels are presented as fold difference relative to the subpopulation of lowest detection and were statistically compared via one-way ANOVA. *P* values were adjusted via the Benjamini-Hochberg method to account for multiple comparisons. Arrowheads indicate the leukocyte subpopulation of highest expression

department admission. The expression levels of each of the seven genes which have been previously identified as being upregulated in whole blood during AIS were strongly positively associated with NMLR across the total patient population. Conversely, the expression levels of all three genes which have been previously identified as being downregulated in whole blood during AIS were strongly negatively associated with NMLR (Fig. 2). Taken as a whole, these relationships with NMLR, along with the strong pattern of lineage-specific expression observed in our analysis of isolated leukocytes, suggest that the whole-blood expression levels of many if not all of these genes likely serve as markers of the cellular composition of the circulating leukocyte pool, and that their differential expression in response to stroke simply reflects a shift in WBC counts.

A Majority of Variance in the Expression Levels of Stroke-Associated Genes Is Attributable to Variance in WBC Counts

We then wanted to quantify what proportion of the variance in whole-blood expression levels across the ten previously identified stroke-associated genes was attributable to variance in WBC count alone; factor analysis was used to extract the structure of latent factors underlying the cumulative variance in expression across the ten genes, and the relationships between the extracted factors and WBC differential was assessed across the total patient population. The whole-blood expression levels of the ten genes exhibited a high degree of intercorrelation. Remarkably, hierarchal clustering of the genes based on the covariance in their whole-blood expression levels produced perfect clustering according to their predominant leukocyte subpopulations of expression observed in our analysis of isolated leukocytes (Fig. 3a), further supporting the idea that the whole-blood expression levels of these genes serve as a marker of WBC counts. The predominant factor extracted in factor analysis (Factor 1) captured 54.3% of the collective variance in whole-blood expression levels across the ten genes (Fig. 3b) and exhibited a robust positive correlation with NMLR (Fig. 3c); taken together, these two pieces of information infer that potentially over half the collective variance in the whole-blood expression levels of the ten genes across the patient population were attributable solely to underlying variance in WBC counts.

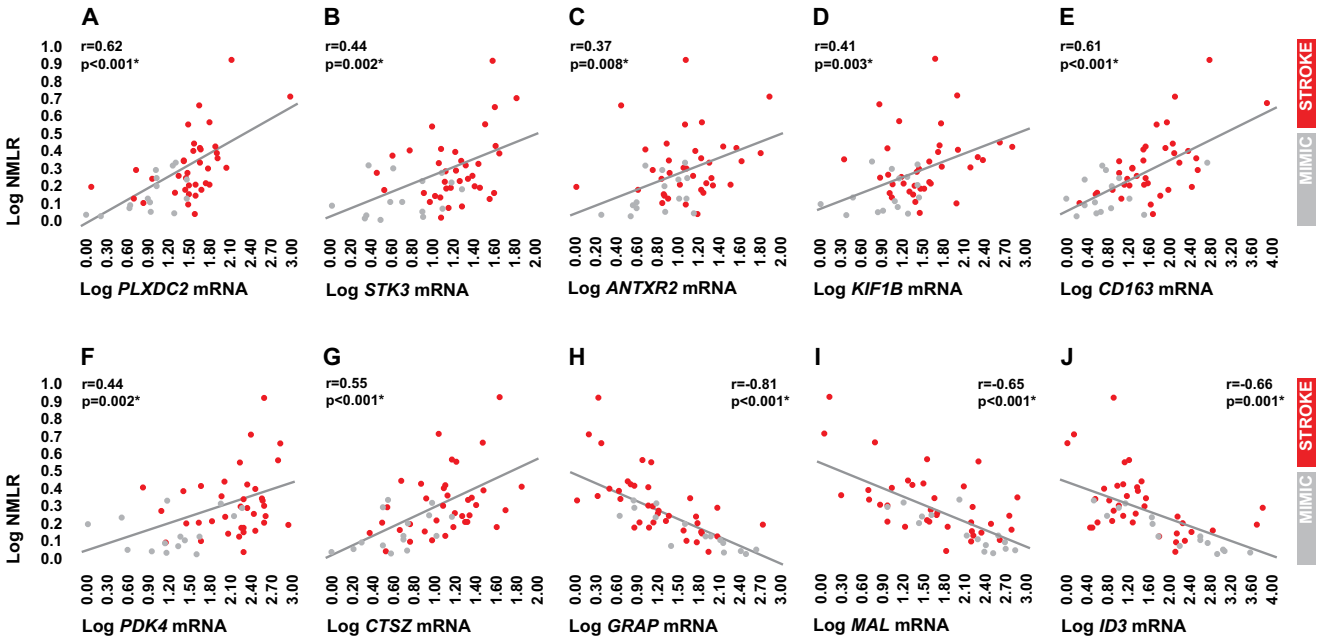

Fig. 2 Relationship between the whole-blood expression levels of stroke-associated genes and white blood cell differential. Relationship between peripheral whole-blood expression levels of the ten stroke-associated genes and NMLR in AIS patients and stroke mimics at emergency department admission. The strength and significance of relationships were testing using Spearman's rho. *P* values were adjusted via the Benjamini-Hochberg method to account for multiple testing. A further comprehensive analysis of the relationships between gene expression levels and absolute cell counts can be found in Supplementary Fig. 1

Stroke-Associated Genes Exhibit Superior Diagnostic Performance Relative to WBC Differential Alone

Lastly, we compared the ability of the whole blood expression levels of the ten previously identified stroke-associated genes to discriminate between AIS patients and stroke mimics to that of WBC alone. We observed an identical pattern of whole-blood differential expression between stroke patients and stroke mimics across the ten genes as described in our prior work (Fig. 4a), and their collective unmanipulated expression levels were able to discriminate between groups with high levels of accuracy using kNN for classification (Fig. 4b). Expectedly, NMLR was significantly elevated in stroke patients in comparison to stroke mimics (Fig. 4a); however, NMLR displayed a statistically inferior ability to differentiate between groups relative to the collective unmanipulated expression levels of the ten stroke-associated genes when comparing area under ROC curve (Fig. 4b). Unsurprisingly, the

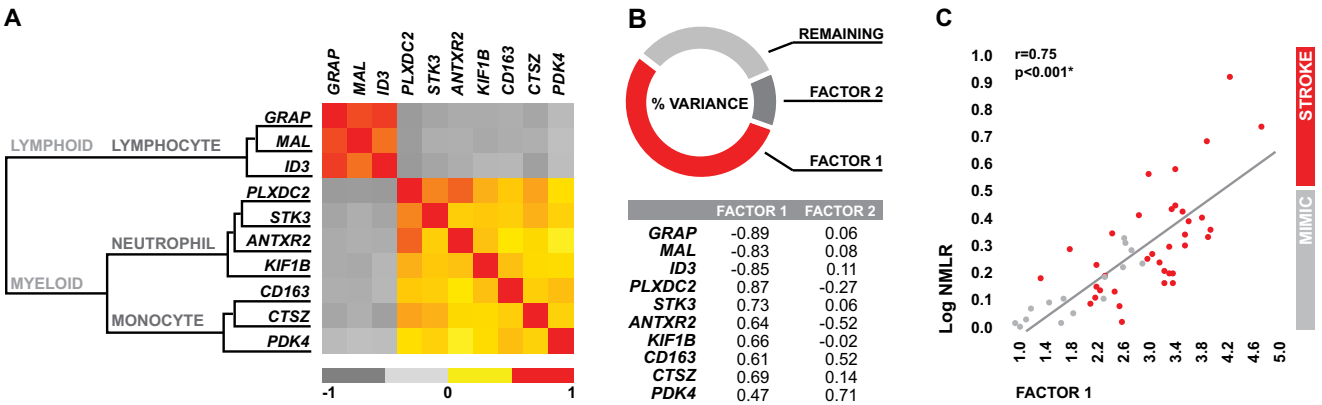

Fig. 3 Analysis of latent factors which drive the expression levels of stroke-associated genes in whole blood. **a** Covariance matrix depicting the correlations between the whole-blood expression levels of the ten stroke-associated genes across the total patient population. Strength of correlations were assessed via Pearson's *r*. Hierarchical clustering indicates similarity in expression based on covariance; the predominant leukocyte subpopulation of expression for each major cluster of genes is indicated in the cluster dendrogram. **b** The proportion of total collective variance in the whole-blood expression levels of the ten stroke-associated genes attributable to the significant latent factors identified via factor analysis, along with the loadings associated with each factor. **c** The relationship between the predominate latent factor identified in factor analysis and NMLR across the total patient population

predominant factor extracted from the whole-blood expression levels of the ten genes in our factor analysis displayed similar diagnostic ability as NMLR (Fig. 4b). Taken together, this suggests that even though a majority of the variance in the collective whole-blood expression levels of ten genes is likely directly driven by WBC differential, there are other factors that influence their whole-blood expression levels that are additive in terms of their diagnostic performance.

Discussion

Our group recently identified a panel of ten genes whose whole-blood expression levels have demonstrated an ability to detect ischemic stroke with high levels of accuracy [3, 6]. The goal of this study was to examine the potential mechanisms by which this ten-marker panel of genes becomes differentially expressed in whole blood in response to stroke. Collectively, our results infer that a substantial degree of the stroke-induced differential expression observed across the ten genes occurs as an artifact of shifts in WBC differential, and not through transcriptional modifications at the level of individual cells. While this does not diminish the potential of these genes for diagnostic use, it suggests that it is unlikely that they are biological drivers of the stroke-induced immune response. These findings not only provide insight regarding this particular panel of ten genes but also into the results of prior genome-wide stroke transcriptomics studies performed in whole blood.

It is well established that stroke induces a rapid rise in circulating counts of myeloid cell populations such as neutrophils and monocytes [10, 11] and a simultaneous drop in counts of lymphoid cell populations such as T- and B-lymphocytes [11]. Due to transcriptional diversity existing across the leukocyte subpopulations which comprise the peripheral immune system [16], such shifts in WBC counts can alter gene expression at the level of whole blood, even in the absence of transcriptional changes at the cellular level [17]. All ten previously identified stroke-associated genes exhibited some degree of expressional heterogeneity across the leukocyte subpopulations assessed in our isolated cell analysis, suggesting that their whole-blood expression levels are sensitive to perturbations in WBC differential. Thus, it was unsurprising to find that the whole-blood expression levels of each of the ten genes was correlated with WBC counts in AIS patients and stroke mimics at emergency department admission. Furthermore, our factor analysis inferred that over 50% of the cumulative variance in whole-blood expression levels across the ten genes in the overall patient population was driven by underlying variance in WBC counts alone. Taken together, our collective results (summarized in Table 2) suggest that a majority of the differential expression observed across the ten genes in response to stroke occurs simply as a marker of the stroke-induced shift in the cellularity of the peripheral immune system.

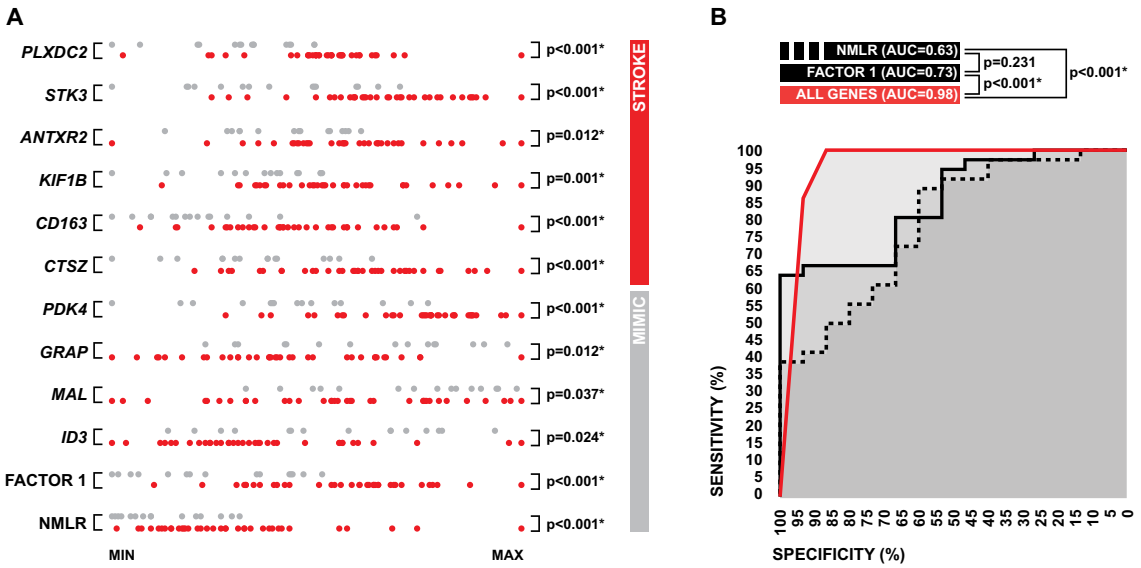

Fig. 4 Diagnostic performance of stroke-associated genes relative to WBC differential. **a** The peripheral whole-blood expression levels of the ten stroke-associated genes, the predominant latent factor identified in factor analysis, along with NMLR, compared between AIS patients and stroke mimics at emergency department admission. All variables were scaled on an interval from zero to one using unity-based normalization. Intergroup comparisons were made via two-sample two-tailed *t* test, and *p* values were adjusted via the Benjamini-Hochberg method to account for multiple comparisons. **b** ROC curve depicting the collective ability of the whole-blood expression levels of the ten stroke-associated genes to discriminate between AIS patients and stroke mimics using kNN ($k = 3$), compared that of the predominant factor identified in factor analysis and NMLR. Areas under ROC curves were statistically compared via the DeLong method, and *p* values were adjusted via the Benjamini-Hochberg method to account for multiple comparisons

Table 2 Summary of findings

Gene	Change in whole blood expression during AIS	Predominant leukocyte subpopulation of expression	Change in circulating leukocyte subpopulation count during AIS[a]	Correlation between whole blood expression and NMLR
PLXDC2	▲	Myeloid (neutrophil)	▲	+
STK3	▲	Myeloid (neutrophil)	▲	+
ANTXR2	▲	Myeloid (neutrophil)	▲	+
KIF1B	▲	Myeloid (neutrophil)	▲	+
CD163	▲	Myeloid (monocyte)	▲	+
CTSZ	▲	Myeloid (monocyte)	▲	+
PDK4	▲	Myeloid (monocyte)	▲	+
GRAP	▼	Lymphoid (B-lymphocyte)	▼	–
MAL	▼	Lymphoid (T-lymphocyte)	▼	–
ID3	▼	Lymphoid (B-lymphocyte)	▼	–

[a] As reported in Vogelgesang et al. [11]

In terms of immunopathology, our findings suggest that a majority of the ten genes evaluated in this study are unlikely to be significant biological drivers of the peripheral immune response to stroke, as the genes which drive the response are likely to be those which exhibit true transcriptional modifications at the cellular level. However, it is relevant to note that there are a limited number of scenarios in which this may not hold true, such as in the case of genes which encode for secreted proteins; the concentrations of secreted protein products may become altered in systemic circulation and exert biological effects simply by virtue of changes in the number of secreting cells, independently of any change in transcription or translation at the level of the individual cells themselves. One of the ten stroke-associated genes investigated in our analysis, *CD163*, encodes for a protein known as cluster of differentiation 163 (CD163) which has the potential for secretion [28]; elevated circulating levels of soluble CD163 have been previously reported in stroke [29], and recent studies have implicated soluble CD163 as being mechanistically involved in multiple aspects of stroke immunopathology [29–31]. However, CD163 likely represents an exception with regard to the ten-gene panel, as a majority of its genes encode for various intracellular proteins (Supplemental Table 2) and it is unlikely that they play a similar mechanistic role in driving the stroke-induced peripheral immune response.

From a broader prospective, our findings highlight the limitation in trying to draw conclusions about the mechanisms which drive stroke immunopathology using gene expression data generated from whole blood. Without additional experimentation, it is impossible to determine which genes are differentially expressed simply as artifacts of the stroke-induced shift in WBC differential, and which are differentially expressed due to true transcriptional modifications at the cellular level; the inability to make this distinction makes the identification of pathologically relevant genes difficult. However, a majority of genome-wide stroke transcriptomics investigations performed in whole blood have interpreted their results under the assumption that all observed differential expressions are a result of true cellular alterations in transcription [1–5]. Our findings clearly infer that this assumption is likely untrue, as it is highly likely that several genes identified as being differentially expressed in these previous studies undergo altered expression via a similar mechanism as the ten which we assessed in our analysis. Exploratory experiments from our laboratory support this notion, as genes such as *ARG1*, *NPL*, and *CCR7* which have been identified as being differentially regulated in these prior investigations [1, 2, 5] display similar expressional heterogeneity within the peripheral immune system, and similar relationships with WBC differential, as the ten genes which we focused on in this study (Supplemental Fig. 2).

Thus, it is likely that some of the conclusions generated from these previous studies regarding the potential mechanisms which drive stroke immunopathology may be flawed. For example, based on observations of decreased expression of *CCR7* in whole blood during stroke, it has been suggested that transcriptional inhibition of chemokine receptor 7 (CCR7) signaling mechanistically drives the reduction lymphocyte counts observed in pathology [5, 32]. However, because *CCR7* exhibits thousand-fold higher expression levels on lymphocytes than on other leukocyte subpopulations, it is more likely that *CCR7* expression is reduced in whole blood simply as an artifact of the stroke-induced drop in lymphocyte counts. While this is just a single example, it is representative of numerous potentially confounded conclusions made regarding stroke immunopathology based on whole-blood transcriptomic data. It is pertinent to note that the potential confounds associated with attempting to draw mechanistic conclusions regarding stroke immunopathology through whole-blood analysis apply not only in the case of coding mRNA as largely discussed here, but also in the case of non-

coding RNA species such as lncRNA and miRNA [33, 34]. Depending on the research question, future studies employing transcriptional profiling of the peripheral immune system to explore stroke immunopathology may want to avoid such confounds by analyzing isolated leukocyte subpopulations, or ideally, single cells.

While our results suggest that it is unlikely that a majority of the ten genes which we assessed in our investigation are biological drivers of stroke immunopathology, it does not diminish their potential clinical usefulness as biomarkers. Despite the fact that a substantial proportion of the collective variance in whole-blood expression levels across the ten genes was likely driven by underlying variance in WBC counts as captured by NMLR, the cumulative expression levels of the ten genes displayed a significantly greater ability to discriminate between AIS patients and stroke mimics relative to NMLR alone. This suggests that there are other factors that influence the whole-blood expression levels of the genes not explainable via NMLR which are additive in terms of their diagnostic performance. Most probable of these factors are stroke-induced changes in WBC counts not captured by NMLR, such as shifts in the ratios of more finite leukocyte populations such as T-lymphocytes and B-lymphocytes within the predominate leukocyte lineages. In the case of some of the genes, these factors could also include modest stroke-induced modifications in cellular transcription which drive a minority of the collective variance across the panel as a whole, but nonetheless impart some diagnostic power.

By providing improved understanding of the mechanisms which drive the differential expression of the stroke-associated genes assessed in our analysis, our findings add to their potential value as a biomarker panel by imparting biological context which will allow for improved clinical interpretability in the scenario that real-world implementation is realized. Most importantly, our findings provide insight regarding situations which are most likely to yield erroneous tests. For example, due to the likelihood that a substantial proportion of the variance in gene expression levels across the panel is driven by WBC counts, patients presenting with infectious disease at the time of stroke evaluation may be more prone to false-positive results. Conversely, patients being evaluated hyper-acutely with regard to time of symptom onset may be more prone to false-negative results, as the shifts in WBC counts induced by stroke become magnified the longer pathology progresses [12]. Additional investigation into the aforementioned scenarios could be useful in shaping future recommendations for clinical use.

It is important to note that this study was not without limitations. For example, our conclusion that stroke-induced shifts in WBC counts substantially drive the differential expression of the investigated genes was based largely on associative data; however, we feel that the overall pattern of associations which we observed across the experiments overwhelmingly supported this interpretation. It could also be argued that performing our analysis of isolated leukocytes using cells harvested from healthy donors as opposed to cells harvested from stroke patients reduced the pathological relevance of our findings; however, we feel it is highly unlikely that the overall transcriptional distributions of the analyzed genes would look dramatically different in stroke patients in comparison to healthy donors, as the patterns of expression which we observed across the leukocyte subpopulations were highly distinct and extremely consistent across all analyzed subjects.

Collectively, our results provide clarity regarding the mechanisms by which the ten previously identified stroke-associated genes explored in our analyses become differentially regulated in whole blood during stroke, information which will prove useful if these genes are to be investigated further for biomarker-based stroke detection. More broadly, our findings provide valuable insight regarding the influence of stroke-induced shifts in leukocyte counts on whole-blood RNA expression, and highlight the potential pitfalls associated with attempting to draw conclusions about the biologic drivers of stroke immunopathology using expression data generated from whole blood.

Acknowledgments The authors would foremost like to thank the subjects and their families, as this work was truly made possible by their selfless contribution. In addition, we would like to thank the stroke team at Ruby Memorial Hospital for supporting this research effort.

Funding Work was funded via a Robert Wood Johnson Foundation Nurse Faculty Scholar award to TLB (70319), a National Institutes of Health CoBRE sub-award to TLB (P20 GM109098), and Case Western Reserve University FPB School of Nursing start-up funds issued to GCO.

Compliance with Ethical Standards

Procedures were approved by the institutional review boards of West Virginia University and Ruby Memorial Hospital (IRB protocol 1410450461R001), and written informed consent was obtained from all subjects or their authorized representatives prior to study procedures.

Conflict of Interest GCO and TLB have a patent pending re: genomic patterns of expression for stroke diagnosis. TLB serves as chief scientific officer for Valtari Bio Incorporated. Work by GCO is part of a pending licensing agreement with Valtari Bio Incorporated. GCO has received consulting fees from Valtari Bio Incorporated. The remaining authors report no potential conflicts of interest.

References

1. Tang Y, Xu H, Du X, Lit L, Walker W, Lu A, et al. Gene expression in blood changes rapidly in neutrophils and monocytes after ischemic stroke in humans: a microarray study. J Cereb Blood Flow Metab Off J Int Soc Cereb Blood Flow Metab. 2006;26:1089–102.
2. Stamova B, Xu H, Jickling G, Bushnell C, Tian Y, Ander BP, et al. Gene expression profiling of blood for the prediction of ischemic stroke. Stroke J Cereb Circ. 2010;41:2171–7.
3. O'Connell GC, Petrone AB, Treadway MB, Tennant CS, Lucke-Wold N, Chantler PD, et al. Machine-learning approach identifies a

pattern of gene expression in peripheral blood that can accurately detect ischaemic stroke. Npj Genomic Med. 2016;1:16038–8.
4. Raman K, O'Donnell MJ, Czlonkowska A, Duarte YC, Lopez-Jaramillo P, Peñaherrera E, et al. Peripheral blood MCEMP1 gene expression as a biomarker for stroke prognosis. Stroke J Cereb Circ. 2016;47:652–8.
5. Barr TL, Conley Y, Ding J, Dillman a, Warach S, Singleton a, et al. Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling. Neurology 2010;75:1009–1014.
6. O'Connell GC, Chantler PD, Barr TL. Stroke-associated pattern of gene expression previously identified by machine-learning is diagnostically robust in an independent patient population. Genomics Data. 2017;14:47–52.
7. Chamorro Á, Meisel A, Planas AM, Urra X, van de Beek D, Veltkamp R. The immunology of acute stroke. Nat Rev Neurol. 2012;8:401–10.
8. Iadecola C, Anrather J. The immunology of stroke: from mechanisms to translation. Nat Med. 2011;17:796–808.
9. Macrez R, Ali C, Toutirais O, Le Mauff B, Defer G, Dirnagl U, et al. Stroke and the immune system: from pathophysiology to new therapeutic strategies. Lancet Neurol. 2011;10:471–80.
10. Urra X, Cervera Á, Obach V, Climent N, Planas AM, Chamorro A. Monocytes are major players in the prognosis and risk of infection after acute stroke. Stroke. 2009;40:1262–8.
11. Vogelgesang A, Grunwald U, Langner S, Jack R, Bröker BM, Kessler C, et al. Analysis of lymphocyte subsets in patients with stroke and their influence on infection after stroke. Stroke. 2008;39:237–41.
12. Guo Z, Yu S, Xiao L, Chen X, Ye R, Zheng P, et al. Dynamic change of neutrophil to lymphocyte ratio and hemorrhagic transformation after thrombolysis in stroke. J Neuroinflammation. 2016;13:199–9.
13. Rosell A, Cuadrado E, Ortega-Aznar A, Hernández-Guillamon M, Lo EH, Montaner J. MMP-9-positive neutrophil infiltration is associated to blood-brain barrier breakdown and basal lamina type IV collagen degradation during hemorrhagic transformation after human ischemic stroke. Stroke. 2008;39:1121–6.
14. Maestrini I, Strbian D, Gautier S, Haapaniemi E, Moulin S, Sairanen T, et al. Higher neutrophil counts before thrombolysis for cerebral ischemia predict worse outcomes. Neurology. 2015;85(16):1408–16.
15. Urra X, Cervera Á, Villamor N, Planas a M, Chamorro Á. Harms and benefits of lymphocyte subpopulations in patients with acute stroke. Neuroscience. 2009;158:1174–83.
16. Abbas a R, Baldwin D, Ma Y, Ouyang W, Gurney a MF, et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. Genes Immun. 2005;6:319–31.
17. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, et al. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci U A. 2003;100:1896–901.
18. Kidwell CS, Warach S. Acute ischemic cerebrovascular syndrome: diagnostic criteria. Stroke. 2003;34:2995–8.
19. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. Lancet Lond Engl. 1991;337:1521–6.
20. Lee LJ, Kidwell CS, Alger J, Starkman S, Saver JL. Impact on stroke subtype diagnosis of early diffusion-weighted magnetic resonance imaging and magnetic resonance angiography. Stroke. 2000;31:1081–9.
21. Heckmann L-H, Sørensen PB, Krogh PH, Sørensen JG. NORMA-gene: a simple and robust method for qPCR normalization based on target gene data. BMC Bioinformatics. 2011;12:250–0.
22. O'Connell GC, Treadway MB, Petrone AB, Tennant CS, Lucke-Wold N, Chantler PD, et al. Leukocyte dynamics influence reference gene stability in whole blood: data-driven qRT-PCR normalization is a robust alternative for measurement of transcriptional biomarkers. Lab Med. 2017;48:346–56.
23. Ross I, Robert G, Ihaka R, Gentleman R. R: a language for data analysis and graphics. J Comput Graph Stat. 1996;5:299–314.
24. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:1–8.
25. Venables WN, Ripley BD. Modern applied statistics with S [Internet]. Springer New York: New York, NY; 2002. Available from: http://www.stats.ox.ac.uk/pub/MASS4
26. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837–45.
27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995:289–300.
28. Droste A, Sorg C, Högger P. Shedding of CD163, a novel regulatory mechanism for a member of the scavenger receptor cysteine-rich family. Biochem Biophys Res Commun. 1999;256:110–3.
29. O'Connell GC, Tennant CS, Lucke-Wold N, Kabbani Y, Tarabishy AR, Chantler PD, et al. Monocyte-lymphocyte cross-communication via soluble CD163 directly links innate immune system activation and adaptive immune system suppression following ischemic stroke. Sci Rep. 2017;7:12940–0.
30. Leclerc JL, Lampert AS, Loyola Amador C, Schlakman B, Vasilopoulos T, Svendsen P, et al. The absence of the CD163 receptor has distinct temporal influences on intracerebral hemorrhage outcomes. J Cereb Blood Flow Metab. 2017;0271678X1770145-0271678X1770145.
31. Liu R, Cao S, Hua Y, Keep RF, Huang Y, Xi G. CD163 expression in neurons after experimental intracerebral hemorrhage. Stroke. 2017;48:1369–75.
32. Petrone AB, O'Connell GC, Regier MD, Chantler PD, Simpkins JW, Barr TL. The role of arginase 1 in post-stroke immunosuppression and ischemic stroke severity. Transl Stroke Res. 2016;7:103–10.
33. Jickling GC, Ander BP, Zhan X, Noblett D, Stamova B, Liu D. MicroRNA expression in peripheral blood cells following acute ischemic stroke and their predicted gene targets. PLoS One. 2014;9:e99283.
34. Dykstra-Aiello C, Jickling GC, Ander BP, Shroff N, Zhan X, Liu D, et al. Altered expression of long noncoding RNAs in blood after ischemic stroke and proximity to putative stroke risk loci. 2016.