



# High-Throughput Profiling of Circulating Antibody Signatures for Stroke Diagnosis Using Small Volumes of Whole Blood

Grant C. O'Connell<sup>1</sup> · Phillip Stafford<sup>2</sup> · Kyle B. Walsh<sup>3,4</sup> · Opeolu Adeoye<sup>3,4</sup> · Taura L. Barr<sup>5</sup>

Published online: 19 February 2019  
© The American Society for Experimental NeuroTherapeutics, Inc. 2019

## Abstract

Accurate stroke recognition during triage can streamline care and afford patients earlier access to life-saving interventions. However, the tools currently available to clinicians for prehospital and early in-hospital identification of stroke are limited. The peripheral immune system is intricately involved in stroke pathology and thus may be targetable for the development of immunodiagnostics. In this preliminary study, we sought to determine whether the circulating antibody pool is altered early in stroke, and whether such alterations could be leveraged for diagnosis. One hundred microliters of peripheral whole blood was sampled from 19 ischemic stroke patients, 17 hemorrhagic stroke patients, and 20 stroke mimics in the acute phase of care. A custom-fabricated high-density peptide array comprising 125,000 unique probes was used to assess the binding characteristics of blood-borne antibodies, and a random forest-based approach was used to select a parsimonious set of probes with an optimal ability to discriminate between groups. The coordinate antibody binding intensities of the top 17 probes identified in our analysis displayed an ability to differentiate the total pool of stroke patients from stroke mimics with 92% sensitivity and 90% specificity, as well as detect hemorrhage with 88% sensitivity and 87% specificity, as determined using a same-set cross-validation. These preliminary findings suggest that stroke-associated alterations in the circulating antibody pool may have clinical utility for diagnosis during triage, and that such a possibility warrants further investigation.

**Key Words** Biomarkers · triage · proteomics · molecular diagnostics · machine-learning · immunology

## Background

Due to the time-efficacy relationship associated with acute stroke interventions, tools which allow for accurate stroke

diagnosis during triage have the potential to streamline care and improve patient outcomes. In the prehospital setting, confident recognition of stroke by emergency medical services (EMS) personnel allows for direct transport to certified stroke centers, which not only saves time, but also affords patients access to advanced treatment options not available at smaller medical facilities. Beyond the initial recognition of stroke, the ability to identify additional pathological features such as stroke subtype allows for advanced notice to be given to the receiving medical center, which can be used to expedite treatment once the patient arrives. In the case of patients arriving to the hospital through non-EMS mechanisms, the ability to make similar diagnostic determinations at the initial point of contact in the emergency department can be equally beneficial by allowing for expedited stroke team referral.

Currently, critical early triage decisions are often made based on evaluations performed by clinicians with relatively little-specialized stroke training using symptom-based severity and recognition scales. The accuracy of these scales in such scenarios is limited, and it is currently estimated that up to

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13311-019-00720-9>) contains supplementary material, which is available to authorized users.

---

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

<sup>1</sup> School of Nursing, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106-4904, USA

<sup>2</sup> Biodesign Institute, Arizona State University, Tempe, Arizona, USA

<sup>3</sup> Department of Emergency Medicine, College of Medicine, University of Cincinnati, Cincinnati, Ohio, USA

<sup>4</sup> Gardner Neuroscience Institute, University of Cincinnati, Cincinnati, Ohio, USA

<sup>5</sup> Valtari Bio Incorporated, Morgantown, West Virginia, USA

30% of stroke patients are misdiagnosed during initial presentation, leading to life-threatening delays in care [1–3]. Because of this, there has been a push for the identification of stroke-associated peripheral blood biomarkers which could be rapidly measured at the point-of-care to better inform early clinical decision-making.

The peripheral immune system plays a central role in stroke pathology; not only is there a rapid systemic inflammatory response to the acute injury [4], but emerging evidence suggests that peripheral immune changes may precede symptom onset and in some cases trigger the acute event itself [5]. Recent studies by our group and others have demonstrated that this phenomenon can be targeted diagnostically; genome-wide transcriptomic profiling of circulating immune cells has identified RNA-based biomarker panels with the ability to accurately identify ischemic stroke and predict various post-stroke complications [6–10]. However, due to technical limitations, RNA-based biomarkers have limited utility in prehospital and emergency care settings; viable methods for accurate measurement of nucleic acids at the point-of-care are lacking, as existing platforms for rapid blood biomarker screening are largely geared towards immunoassay-based protein detection. Unfortunately, prior stroke proteomics investigations have been relatively small in scope, often examining only a small number of predetermined targets, and have produced few candidate protein biomarkers with clinically useful levels of diagnostic accuracy [11].

However, recent advances in high-throughput protein array technologies may allow for broader scale proteomic analyses in stroke, and open the door for the discovery of translationally relevant biomarkers which would be well suited for point-of-care detection. In the cancer field, recently developed high-density non-native peptide arrays which allow for large-scale profiling of the circulating antibody pool are being increasingly used to identify diagnostically robust biomarker panels [12, 13]. This methodology, often referred to as immunosignaturing, utilizes microarrays comprising hundreds of thousands of unique pseudo-random peptide probes ranging from 10 to 20 amino acids in length [14]. Arrays are exposed to patient serum or soluble whole blood fractions to allow for peptide-antibody interaction, rinsed, and bound antibodies are subsequently detected using fluorescently labeled pan antibody against human immunoglobulin. The overall pattern of binding across the array is then analyzed, and binding patterns can ultimately be compared between disease states to identify features which may have utility for diagnosis (Supplementary Fig. 1).

Given the known involvement of the peripheral immune system in stroke pathology, it is possible that stroke induces changes to the circulating antibody pool which could be targeted diagnostically. In this study, we used a custom-fabricated high-density peptide array to investigate the feasibility of antibody-based stroke

diagnosis within an experimental context relevant to triage.

## Materials and Methods

### Experimental Design

A cohort of acute ischemic stroke patients ( $n = 19$ ), hemorrhagic stroke patients ( $n = 17$ ), and stroke mimics ( $n = 20$ ) were recruited at an urban tertiary care medical center (University of Cincinnati Medical Center, Cincinnati, OH). One hundred microliters of peripheral whole blood was sampled at either emergency department admission or upon inter-hospital transfer. Circulating antibody profiles were generated from whole blood samples using a custom-fabricated peptide array, and a two-step machine-learning approach was subsequently used to select peptide probes suitable for stroke diagnosis. First, random forest was used to rank all probes by importance in terms of their ability to discriminate between ischemic stroke, hemorrhagic stroke, and stroke mimic samples. Then recursive feature selection was used to identify the minimum number of top-ranked probes which could provide optimal discriminatory performance. In order to evaluate the robustness of our analysis in terms of its ability to select optimally discriminatory probes, a permutation analysis was performed in which the diagnostic ability of the top-ranked probes was compared to that of probes selected at random.

### Patients

All ischemic and hemorrhagic stroke patients presented with clinical symptoms consistent with acute stroke and displayed definitive radiographic evidence of ischemic or hemorrhagic pathology on magnetic resonance imaging (MRI) or computed tomography (CT). Patients admitted or referred as suspected strokes based on the overt presentation of stroke-like symptoms, but receiving a definitive negative diagnosis for stroke upon neuroradiological imaging and clinical evaluation was identified as acute stroke mimics [15]; the final discharge diagnoses of the stroke mimic group can be found in Supplementary Table 1. All diagnoses were adjudicated by an experienced stroke physician. Patients were excluded if they received a non-definitive diagnosis, were diagnosed with transient ischemic attack (TIA), reported a prior hospitalization within 30 days, were under 18 years of age, or were more than 12 h past symptom onset. Time from symptom onset was determined by the time the patient was last known to be free of neurological symptoms. Injury severity was determined according to the National Institutes of Health Stroke Scale (NIHSS) at the time of blood draw. Demographic information was collected from either the subject or a significant other by a trained clinician.

## Blood Collection

Peripheral whole blood samples were obtained by venipuncture and collected via K<sub>2</sub>EDTA vacutainer. EDTA-treated blood was aliquoted and stored immediately at  $-80\text{ }^{\circ}\text{C}$  until analysis.

## Peptide Array

Silicon wafer peptide arrays, each comprising 125,000 unique probes, were synthesized via photolithography as described previously [14]. Probes were 12 amino acids long and composed of pseudo-random combinations of the 20 natural amino acids, excluding threonine, methionine, isoleucine, and cysteine. Arrays were blocked with 1 mM phosphate-buffered saline containing 3% bovine serum albumin, 0.05% Tween 20 (PBST), and 0.014% mercaptohexanol for 1 h at 25  $^{\circ}\text{C}$  in a darkened humidified chamber prior to use.

One hundred microliters of whole blood was thawed, diluted 1:1 in glycerol to stabilize antibodies, and held at  $-20\text{ }^{\circ}\text{C}$  during processing. Hemocytes and debris were sedimented via centrifugation, and supernatants were collected and diluted 1:1000 in PBST. Diluted soluble whole blood fractions were then incubated on arrays for one hour at 37  $^{\circ}\text{C}$  with gentle agitation to allow for antibody binding. Following incubation, arrays were washed three times for five minutes each with 1 mM tris-buffered saline containing 0.05% Tween 20 (TBST). Arrays were then incubated with 4 nM AlexaFluor 647-conjugated pan anti IgG antibody in PBST for 1 h at room temperature with gentle agitation. Following incubation, arrays were washed three times for 5 min each with TBST, followed by three washes with distilled water. Arrays were then dried by centrifugation and scanned by an Innopsys (Carbonne, France) Innoscan 910 0.5-um color scanner. Images were stored as 16-bit uncompressed TIFFs and aligned using GenePix Pro 6.0 (Molecular Devices, Santa Clara, CA). Raw probe intensities were quantile normalized via the `normalize.quantiles()` function of the “preprocessCore” package for R (R project for statistical computing) [16].

## Random Forest

Machine learning-based analysis of peptide array data was performed using a similar approach as described previously [17]. Random forest models were generated via the “randomForest” package for R [18]. Representative decision trees associated with random forest models were selected and visualized using the “reptree” package [19].

For ranking of probe importance, five replicate random forest models were built, each discriminating between ischemic stroke, hemorrhagic stroke, and stroke mimic samples using the  $\log_2$  transformed normalized intensity values of all 125,000 probes as input. 1.5 million decision trees were

generated for each model, and probe importance was assessed in terms of node purity metrics, as quantified by mean decrease Gini coefficient. Mean decrease Gini coefficient for each probe was averaged across all five models and subsequently used for ranking. The script used for assessment of probe importance is depicted in Supplementary Fig. 2.

For recursive feature selection, successive combinations of the top-ranked probes were evaluated for their ability to discriminate between experimental groups using random forest starting with the top probe, and sequentially preceding through the top 1000. Models were built using 50 times the number of decision trees relative to the number of input probes. For each random forest model, cross-validation prediction probabilities were generated according to out-of-bag vote distributions, yielding a predicted probability of ischemic stroke, hemorrhagic stroke, and stroke mimic for each sample. Hemorrhagic stroke and ischemic stroke prediction probabilities were combined to produce a total stroke prediction probability. Total stroke prediction probability was used to classify samples as stroke/no stroke, and hemorrhagic stroke prediction probabilities were used directly to classify samples as hemorrhage/no hemorrhage (Fig. 1). Model classifications were then compared to true clinical diagnoses to assess accuracy. The script used to generate prediction probabilities for recursive feature selection is depicted in Supplementary Fig. 3.

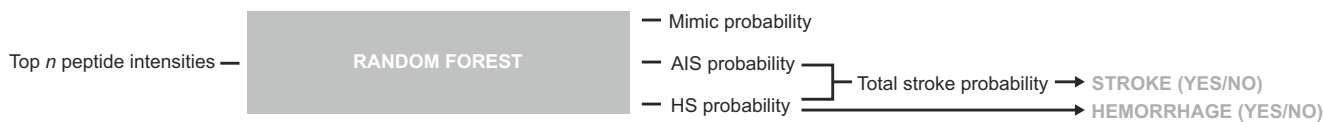
For permutation analysis, 100 unique combinations of  $n$  probes were selected from the total probe pool, and the average diagnostic accuracy across the combinations was compared to that of the top  $n$ -ranked probes. For example, the diagnostic accuracy of the top 10 probes was compared to the average diagnostic accuracy of 100 combinations of 10 randomly selected probes. Random probe combinations were generated using the R `sample()` function. The script used to generate prediction probabilities for permutation analysis is depicted in Supplementary Fig. 4.

## Analysis of Potentially Confounding Factors

In order to determine whether antibody binding signatures selected in random forest were influenced by intergroup differences in clinical and demographic characteristics, random forest prediction probabilities were regressed against clinical diagnoses along with potentially confounding covariates using the `lm()` function of the R base package. Variance decomposition was subsequently employed to determine the relative influence of each covariate on prediction probabilities using the “relaimpo” package [20] as described previously [6].

## Statistics

All statistics were performed using R version 3.4. The level of significance was established at 0.05 for all statistical testing.



**Fig. 1** Paradigm used for sample classification. Cross-validation prediction probabilities were generated from random forest models according to out-of-bag vote distributions, yielding a predicted probability of ischemic stroke (IS), hemorrhagic stroke (HS), and stroke mimic for each sample. Hemorrhagic stroke and ischemic stroke prediction probabilities were

Fisher's exact test was used for comparison of dichotomous variables. Student's *t* test or one-way ANOVA was used for comparison of continuous variables in which appropriate. Hierarchical clustering was performed using the `get_dist()` function of the “factoextra” package, in combination with the `hclust()` function of the R base package. The performance of binary classifiers was assessed via receiver operator characteristic analysis (ROC) via the “pROC” package [21]. The level of significance was established at 0.05 for all statistical testing. *P* values were adjusted using Benjamini-Hochberg method to account for multiple comparisons [22]. Parameters of all statistical tests performed are outlined in detail within the figure legends.

## Results

### Clinical and Demographic Characteristics

The hemorrhagic stroke group included 15 final clinical diagnoses of subarachnoid hemorrhage and 2 final clinical diagnoses of intracerebral hemorrhage. Final clinical diagnoses of the stroke mimic group can be found in Supplementary Table 1. Both ischemic and hemorrhagic stroke patients were significantly older than stroke mimics. Furthermore, ischemic and hemorrhagic stroke patients displayed a greater history of cardiovascular disease than stroke mimics, and a higher prevalence of cardiovascular disease risk factors, especially dyslipidemia. Ischemic stroke and hemorrhagic patients were relatively similar in terms of clinical and demographic characteristics; however, the ischemic stroke group displayed a higher prevalence of dyslipidemia and contained a higher proportion of female subjects (Table 1).

### Model Selection and Diagnostic Performance

The top-ranked peptide probes, as determined by mean decrease Gini coefficient, are depicted in Fig. 2A. The combined ability of the top-ranked probes to differentiate the total pool of ischemic and hemorrhagic stroke patients from stroke mimics in cross-validation is depicted in Fig. 2B. The combined ability of the top-ranked probes to differentiate hemorrhagic stroke patients from ischemic stroke patients and stroke mimics in cross-validation is depicted in Fig. 2C. The top-

ranked probes displayed a markedly better discriminatory ability with regard to both identification of stroke and detection of hemorrhage relative to probes selected at random, suggesting that our analysis was successful in terms of selecting probes with optimal diagnostic characteristics. With respect to both comparisons, cross-validation accuracy appeared to plateau at 17 probes, and thus the model including the top-ranked 17 probes was selected as the final model.

The top 17 probes used in combination were able to discriminate the total pool of ischemic and hemorrhagic stroke patients from stroke mimics with 91.7% sensitivity (95% CI = 77.5–98.2%) and 90.0% specificity (95% CI = 68.3–89.8%, Fig. 3A). The same 17 probes were able to detect hemorrhage with 88.2% sensitivity (95% CI = 63.6–98.5%) and 87.1% specificity (95% CI = 72.7–95.7%) when considering all subjects (Fig. 3B), and 93.3% sensitivity (95% CI = 68.1–99.8%) and 90.0% specificity (95% CI = 68.3–98.8%) when only considering patients first classified as stroke (Fig. 3C).

### Differential Antibody Binding of Top-ranked Probes

A comparison of the antibody binding intensities across the top 17 probes between ischemic stroke patients, hemorrhagic stroke patients, and stroke mimics is shown in Fig. 4A. Significant differences in antibody binding intensities were observed between groups with regard to each of the top 17 probes after controlling for multiple comparisons with the exception of one. Hierarchical clustering of the top 17 probes based on the correlational relationship between their antibody binding intensity levels produced three predominant clusters: one which displayed higher binding intensities in ischemic stroke patients relative to hemorrhagic stroke patients and stroke mimics, one which displayed lower binding intensities in ischemic stroke patients relative to hemorrhagic stroke patients and stroke mimics, and one which displayed higher binding intensities in ischemic and hemorrhagic stroke patients relative to stroke mimics.

Visualization of the final model's most representative decision tree revealed logical node splitting in terms of both the probe importance rankings generated in our probe selection paradigm, as well as the differential antibody binding intensities observed across the top-ranked peptides. For example, the root node of the tree was a split determined by the binding intensity of the top-



**Table 1** Clinical and demographic characteristics

	Group			P value			
	Stroke mimic (n = 20, SM)	Ischemic stroke (n = 19, IS)	Hemorrhagic stroke (n = 17, HS)	Main test	SM v IS	SM v HS	IS v HS
Age <sup>a</sup> (mean ± SD)	57.6 ± 13.9	72.2 ± 14.2	70.4 ± 12.4	0.002*	0.002*	0.005*	0.996
Female <sup>b</sup> , n(%)	12 (60)	10 (52.6)	3 (17.6)	0.023*	0.751	0.017*	0.041*
Caucasian <sup>b</sup> , n(%)	13 (65)	14 (73.7)	13 (76.5)	0.756	–	–	–
African American <sup>b</sup> , n(%)	7 (35)	5 (26.3)	4 (23.5)	0.756	–	–	–
NIHSS <sup>a</sup> (mean ± SD)	1.5 ± 2	11.3 ± 6.8	12.3 ± 9.2	< 0.001*	< 0.001*	< 0.001*	0.683
Minutes to blood draw <sup>a</sup> (mean ± SD)	348.8 ± 155.8	425.6 ± 236.6	508.2 ± 158.7	0.008*	0.096	< 0.001*	0.223
History of stroke <sup>b</sup> , n(%)	0 (0)	6 (31.6)	2 (11.8)	0.009*	0.008*	0.204	0.235
History of myocardial infarction, n(%)	1 (5)	8 (42.1)	2 (11.8)	0.013*	0.008*	0.584	0.065
History of atrial fibrillation <sup>b</sup> , n(%)	3 (15)	8 (42.1)	2 (11.8)	0.069	–	–	–
Hypertension <sup>b</sup> , n(%)	9 (45)	15 (78.9)	11 (64.7)	0.089	–	–	–
Dyslipidemia <sup>b</sup> , n(%)	3 (15)	13 (68.4)	5 (29.4)	0.002*	0.001*	0.428	0.043*
Diabetes <sup>b</sup> , n(%)	5 (25)	4 (21.1)	3 (17.6)	0.920	–	–	–
Current smoking <sup>b</sup> , n(%)	6 (30)	7 (36.8)	2 (11.8)	0.214	–	–	–

<sup>a</sup> Means compared via one-way ANOVA with subsequent planned group-wise comparisons using two-sample two-tailed t-test;

<sup>b</sup> Proportions compared via 2 × 3 Fisher's exact test with subsequent planned group-wise comparisons using 2 × 2 Fisher's exact test; SD, standard deviation;

\* statistically significant.

ranked probe, NVAQAQDENLAG, which displayed lower binding intensities in ischemic stroke patients relative to hemorrhagic stroke patients and stroke mimics. Consistent with this pattern of differential binding, splitting the subject pool based on the root node criterion produced a relatively pure daughter node comprised almost exclusively of ischemic stroke patients, and another daughter node comprised predominantly of hemorrhagic stroke patients and stroke mimics (Fig. 4B).

### Influence of Potentially Confounding Factors on Diagnostic Signature

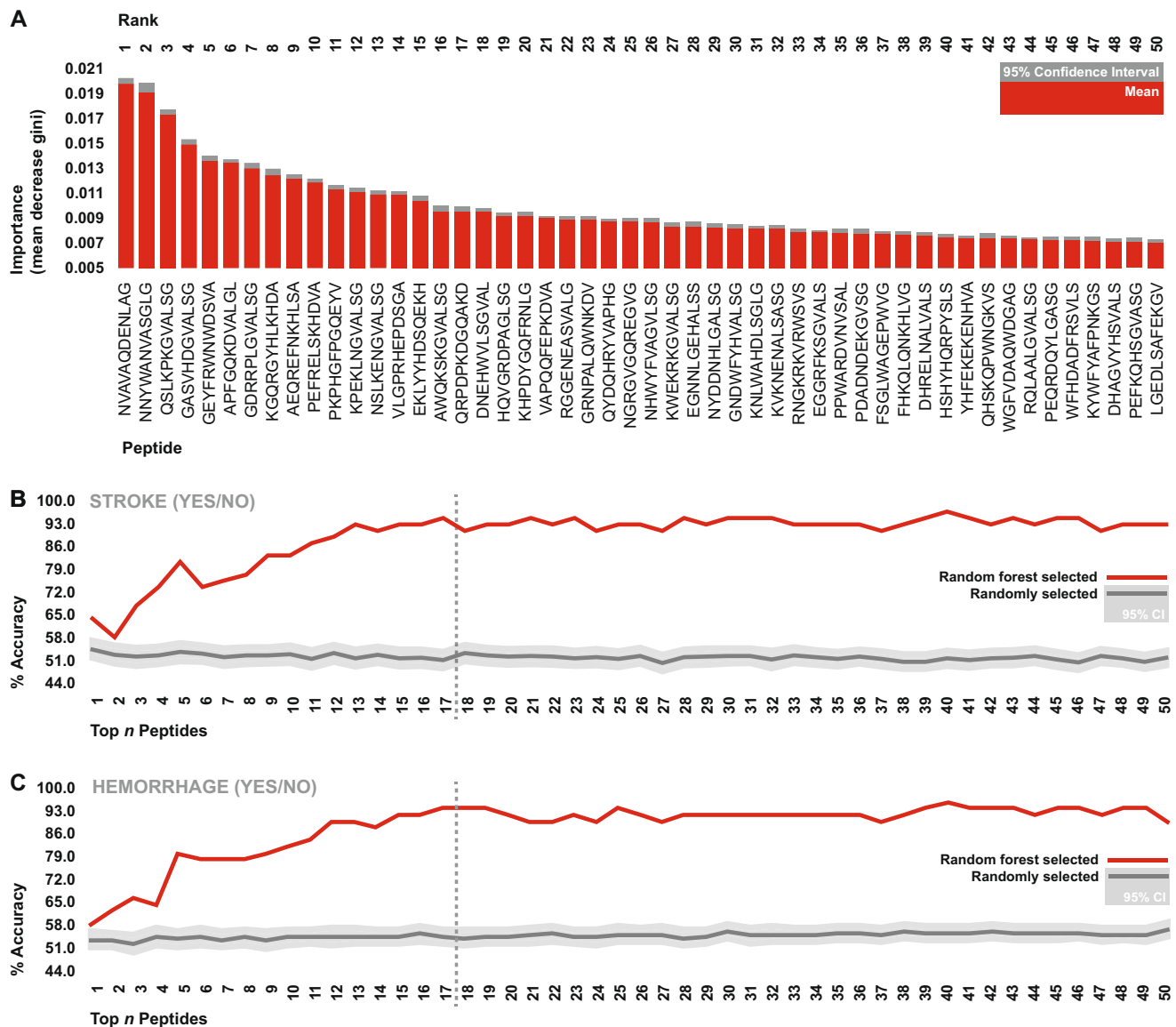
Combined ischemic and hemorrhagic stroke prediction probabilities generated by the final 17-probe random forest model were significantly associated with a clinical diagnosis of stroke independently of age, sex, time to blood draw, history of stroke, history of myocardial infarction, and dyslipidemia in multiple regression. Although stroke prediction probability was also significantly associated with sex, variance decomposition revealed that this association only accounted for a minimal amount of explained variance (Fig. 5A). In addition, hemorrhage prediction probabilities generated by the random forest model were significantly associated with clinical diagnosis of hemorrhage independently of both sex and dyslipidemia when considering the total subject pool, and exhibited no significant association with either potentially confounding factor (Fig. 5B). Collectively, this suggests that the diagnostic

signature identified in our analysis was not largely influenced by intergroup differences in clinical and demographic characteristics.

## Discussion

Accurate stroke recognition during triage can streamline care and afford patients earlier access to life-saving interventions. Unfortunately, the tools currently available to clinicians for identification of stroke in the prehospital and early in-hospital setting are limited. The purpose of this study was to investigate the feasibility of antibody-based stroke diagnosis within an experimental context relevant to triage. Our results provide preliminary evidence that the circulating antibody pool is altered in stroke, and that these alterations can be targeted diagnostically. Albeit in a limited sample size, the antibody-based signature identified in this analysis displayed levels of diagnostic performance which exceed those reported in a majority of prior stroke biomarker investigations, as well as those commonly reported as being achievable using the stroke recognition and severity scales which are currently used in prehospital and emergency medicine settings.

Although it is important to note that no direct comparisons were made, few prior studies have reported similar levels of diagnostic accuracy to those which we observed in this investigation in terms of discriminating between stroke patients and true clinical stroke mimics. Most notably, Dambinova et al.



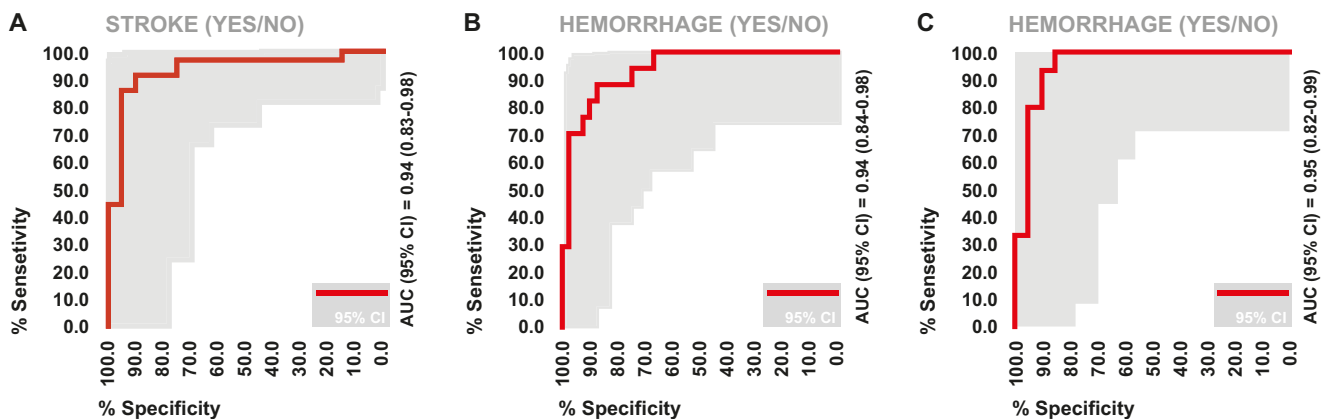
**Fig. 2** Selection of top-ranked probes. (A) Top-ranked probes, ordered by mean decrease Gini coefficient, averaged across five independent random forest models. (B) Combined ability of the antibody binding intensities of the top 50-ranked probes to discriminate the total pool of ischemic and hemorrhagic stroke patients from stroke mimics using random forest,

compared to those of probes selected at random. (C) Combined ability of the antibody binding intensities of the top 50-ranked probes to identify hemorrhagic stroke patients using random forest when considering all subjects, compared to those of probes selected at random

recently reported that plasma levels of brain-derived NR2 peptide could be used to differentiate between stroke patients and a combination of acute stroke mimics and neurologically asymptomatic controls with 93% accuracy [23]. However, a majority of blood samples in this prior study were obtained between 24 and 72 h post symptom onset, and it is currently unknown if NR2 peptide would exhibit an equivalent level of diagnostic performance early in the acute phase of care. The antibody-based signature identified in our analysis was tested earlier in the progression of acute pathology and thus exhibits an obvious advantage in that it has the potential to provide actionable diagnostic information at an early enough time point to influence acute care decisions. With respect to

detection of hemorrhage, a small number of biomarker panels comprising circulating proteins have been previously described which have been reported to discriminate between hemorrhagic stroke patients and ischemic stroke patients with levels of accuracy approaching 90% [24, 25]; however, none of them have displayed the ability to simultaneously differentiate between stroke patients and clinical stroke mimics, which diminishes their value for use as a stand-alone diagnostic test.

In addition to displaying robust performance relative to candidate stroke biomarkers reported in prior investigations, the antibody-based signature identified in our analysis exhibited levels of diagnostic accuracy which exceed those commonly reported with respect to the tools currently being used

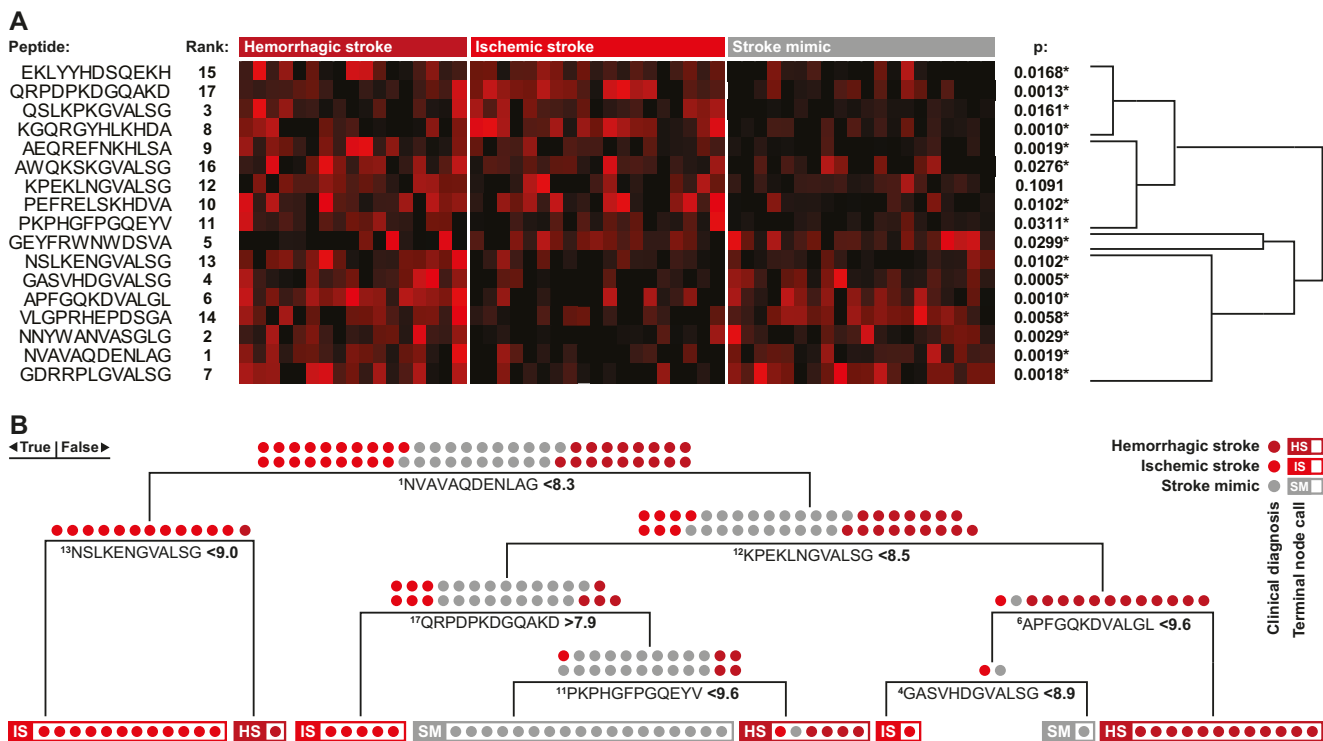


**Fig. 3** Diagnostic ability of the top 17 probes. (A) ROC curve depicting the combined ability of the antibody binding intensities of the top 17-ranked probes to discriminate the total pool of ischemic and hemorrhagic stroke patients from stroke mimics using random forest. (B) Combined ability of the antibody binding intensities of the top 17-ranked probes to

identify hemorrhagic stroke patients using random forest when considering all subjects. (C) Combined ability of the antibody binding intensities of the top 17-ranked probes to identify hemorrhagic stroke patients using random forest when only considering subjects first classified as stroke

guide early triage decisions in clinical practice. A majority of early transport, transfer, and referral decisions are dependent on stroke recognition using rudimentary symptom-based assessments such as the National Institutes of Health Stroke Scale (NIHSS) and the Cincinnati Prehospital Stroke Scale (CPSS). In the hospital setting, the accuracy of such assessments has been reported to range between 58 and 85%

depending on the type of clinician and their level of experience [26]. In the prehospital setting, these assessments are even less effective, in which some studies have documented levels of accuracy as low as 51% [27]. Although it is important to note that no direct comparisons were made, the antibody-based signature identified in our analysis displayed levels of accuracy which, if validated, would constitute a



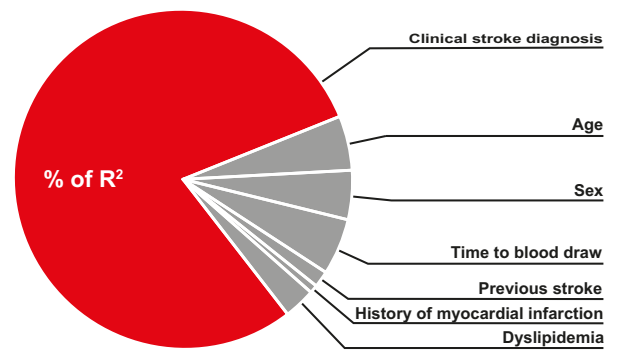
**Fig. 4** Differential antibody binding across the top 17 probes. (A) Antibody binding intensities of the top 17 probes associated with samples from ischemic stroke patients, hemorrhagic stroke patients, and stroke mimics. Binding intensities were statistically compared using one-way ANOVA and *P* values were corrected for multiple comparisons using the Benjamini-Hochberg method. Probes were hierarchically clustered by

similarity in binding intensities as assessed by Spearman's rho. (B) Classification of each subject in the total patient pool according to the final random forest model's most representative decision tree. Each dot represents a single subject. Superscript labels on probes indicate importance ranking

A

Model:  $R^2=0.668$   $p<0.001^*$ 

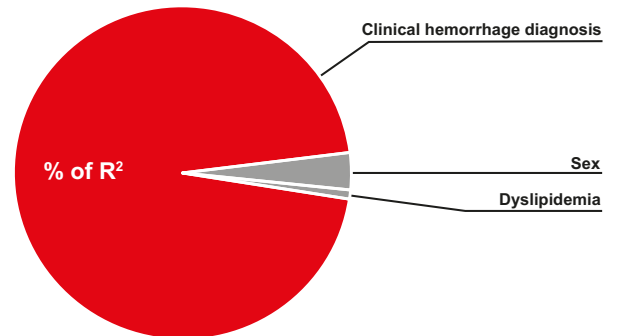
	B	Std error	p value	$R^2$ contribution
Intercept	0.615	0.135	<0.001*	-
Clinical stroke diagnosis	0.514	0.061	<0.001*	0.531 (79.4%)
Age	-0.003	0.001	0.145	0.034 (5.1%)
Sex	-0.139	0.046	0.004*	0.031 (4.7%)
Time to blood draw	-0.001	0.001	0.853	0.036 (5.4%)
Previous stroke	-0.086	0.063	0.182	0.010 (1.6%)
History of myocardial infarction	-0.015	0.065	0.814	0.005 (0.7%)
Dyslipidemia	-0.048	0.053	0.376	0.020 (3.0%)



B

Model:  $R^2=0.563$   $p<0.001^*$ 

	B	Std error	p value	$R^2$ contribution
Intercept	0.231	0.035	<0.001*	-
Clinical hemorrhage diagnosis	0.372	0.045	<0.001*	0.541 (96.1%)
Sex	-0.083	0.042	0.054	0.020 (3.6%)
Dyslipidemia	0.022	0.041	0.589	0.002 (0.3%)



**Fig. 5** Relationships between potentially confounding factors and diagnostic signature. **(A)** Multiple regression model generated by regressing combined ischemic and hemorrhagic stroke prediction probabilities produced by the final 17 probe random forest model against clinical stroke diagnosis and potentially confounding clinical and demographic characteristics. **(B)** Multiple regression model

generated by regressing the hemorrhage prediction probabilities produced by the final 17 probe random forest model against hemorrhage clinical diagnosis and potentially confounding clinical and demographic characteristics. The relative influence of covariates on prediction probabilities as determined by explained variance are graphically depicted

substantial improvement over these symptom-based recognition tools. This suggests that antibody-based stroke diagnostics may have true utility for clinical use if future larger-scale investigations show similar results.

In terms of the biological significance of our findings, our results suggest that the circulating antibody pool is altered in stroke; however, the timing and mechanism of these alterations are unclear. The scalability and unbiased nature of non-native peptide arrays make them well suited for use as diagnostic tools, but because they are not comprised of biologically occurring proteins, it is difficult to elucidate the true antigens associated with differential binding. Previous studies have suggested that stroke-induced blood–brain barrier disruption exposes the peripheral adaptive immune system to novel central nervous antigens [28], which can lead to the production of self-reactive antibodies. Increased circulating levels of antibodies with strong binding affinity towards neural antigens such as myelin basic protein and NMDA receptor have been reported in both human and experimental stroke, however typically days after symptom onset [29–31]. Due to the time it takes the adaptive immune system to produce mature humoral responses [32], it is unlikely that the alterations to the circulating antibody pool which we observed early in acute pathology are a result of a fully formed response against

such antigens; however, they may be a result of the early antibody selection process involving the production of non-homogeneous suboptimal affinity immunoglobulins.

Alternatively, it is possible that the circulating antibody pool is altered prior to acute symptom onset; such alterations could either occur in the immediate days or weeks leading up to the acute injury due to stroke-triggering inflammatory events, or could be the result of long-term immune responses which have developed against precipitating chronic risk factors. A multitude of acute and chronic inflammatory states including infection [33], vasculitis [5], antiphospholipid syndrome [34], rheumatoid arthritis [35], and inflammatory bowel disease [36] are associated with increased risk of stroke. These inflammatory states all involve the adaptive immune system, and can promote thrombosis, alter vascular reactivity, and advance atherosclerosis, which are the direct pathogenic drivers of stroke [5, 37]. Furthermore, chronic risk factors for stroke such as dyslipidemia and atherosclerosis can result in long-term alterations to the complexion of the adaptive immune system. If the diagnostic signature which we measured is in fact already present prior to the acute event due to immune responses which have developed against underlying risk factors as opposed to occurring as a result of an acute response to the injury, it would explain the ability of the profile to discriminate between ischemic and



hemorrhagic stroke, as both pathologies are associated with different risk factor clusters.

In terms the potential for future clinical implementation, the biomarker discovery paradigm which we employed in this investigation included several innovative aspects which make it an optimal starting point for translational development. Prior biomarker discovery studies in stroke have almost exclusively utilized control groups consisting of healthy controls or neurologically normal controls with cardiovascular disease risk factors, and few have sampled blood at ultra-early timepoints [11]. Unsurprisingly, candidate biomarkers identified in these previous studies have repeatedly failed to demonstrate useful levels of diagnostic accuracy when subsequently tested in clinically relevant acute care scenarios. In this investigation, we utilized a control group consisting of true stroke mimics, and sampled blood at emergency department admission, increasing the odds that our findings will generalize in terms of real-world clinical populations.

Furthermore, our use of peptide array as a discovery platform offers additional translational advantages from an assay development perspective. Several commercially available diagnostic devices currently support rapid immunoassay-based blood testing; because a peptide array essentially consists of thousands of parallel sandwich immunoassays consisting of a capture peptide and detection antibody, it should be possible to extract the individual immunoassays deemed most diagnostically useful and develop them for point-of-care detection using existing hardware or established technological principles. Our results suggest that a relatively small number of assays, as few as 17, may be suitable for stroke diagnosis; although the most commonly used point-of-care devices currently found in emergency medicine settings do not support this level of multiplex detection, newer next-generation devices have demonstrated the ability to perform as many as 90 assays in parallel [38]. Thus, it is realistic to think that if a blood-borne antibody-based diagnostic profile for stroke can be validated, it could be developed into a tool ready for clinical use fairly quickly using existing technologies.

It is important to note that this study was not without limitations, many of which were attributable to the preliminary nature of our analysis. Because of our limited sample size, we had to assess the diagnostic performance of our model via same-set cross-validation using out-of-bag error, as opposed to using a more ideal paradigm implementing independent training and validation cohorts. However, due to the use of bootstrap agglomeration, random forest is relatively resistant to overfitting, and out-of-bag error estimates tend to be good predictors of future performance given that the training sample is representative of the overall population of interest [39]. In addition, the small sample size drove our choice to exclude patients with a final diagnosis of TIA; because TIA diagnoses can be ambiguous, and the biological differences between TIA and ischemic stroke are less clearly defined, a much larger subject pool would likely

be needed to generate a model which would could reliably make such distinctions. In spite of these limitations, the collective results reported here are exciting, and a future larger-scale investigation capable of addressing such shortcomings is warranted.

Taken as a whole, our results provide preliminary evidence that there are alterations to circulating antibody pool associated with stroke, and that these alterations can be targeted diagnostically during triage. The antibody-based signature identified in this analysis displayed levels of diagnostic performance which well exceed those reported in previous stroke biomarker investigations, as well as those commonly reported as being achievable via the stroke recognition and severity scales which are currently used in triage scenarios. These findings, taken with the fact that blood-borne antibodies are amenable to point-of-care detection, imply that antibody-based stroke diagnostics could be developed into clinically viable tools with the potential to inform critical decision-making in the acute phase of care.

**Acknowledgments** We would foremost like to thank the patients and their families, as this work was made possible by their selfless contribution. We would also like to thank both the stroke and emergency medicine teams at University of Cincinnati Medical Center for their support.

**Required Author Forms** **Disclosure forms** provided by the authors are available with the online version of this article.

**Author Contributions** Work was conceptualized by GCO. GCO and PS collected molecular data. KBW, OA, and TLB managed subject enrolment and collected clinical data. Statistical analysis was performed by GCO. GCO wrote the manuscript with contributions from PS, KBW, OA, and TLB.

**Funding** This work was funded by Valtari Bio Incorporated.

## Compliance with Ethical Standards

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (detailed in the *Materials and Methods* section) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Conflict of Interest** TLB serves as Chief Scientific Officer of Valtari Bio Incorporated. GCO and TLB have a pending patent re: computer implemented discovery of antibody signatures, as well as a pending patent re: genomic patterns of expression for stroke diagnosis. GCO holds stock in Valtari Bio Incorporated. GCO has received consulting fees from Valtari Bio Incorporated. The remaining authors have no potential conflicts of interest to declare.

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

## References

1. Jiang B, Ru X, Sun H, et al. Pre-hospital delay and its associated factors in first-ever stroke registered in communities from three

- cities in China. *Sci Rep* [Internet]. 2016 Sep [cited 2018 May 3];6(1). Available from: <http://www.nature.com/articles/srep29795>.
2. Lever NM, Nyström KV, Schindler JL, Halliday J, Wira C, Funk M. Missed opportunities for recognition of ischemic stroke in the emergency department. *J Emerg Nurs* 2013;39(5):434–9.
  3. Arch AE, Weisman DC, Coca S, Nyström KV, Wira CR, Schindler JL. Missed ischemic stroke diagnosis in the emergency department by emergency medicine and neurology services. *Stroke* 2016;47(3):668–673.
  4. Iadecola C, Anrather J. The immunology of stroke: from mechanisms to translation. *Nat Med* 2011;17(7):796–808.
  5. Macrez R, Ali C, Toutirais O, et al. Stroke and the immune system: from pathophysiology to new therapeutic strategies. *Lancet Neurol* 2011;10(5):471–480.
  6. O'Connell GC, Petrone AB, Treadway MB, 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(April):16038–16038.
  7. 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.
  8. Stamova B, Xu H, Jickling G, et al. Gene expression profiling of blood for the prediction of ischemic stroke. *Stroke J Cereb Circ* 2010;41(10):2171–7.
  9. Jickling GC, Ander BP, Stamova B, et al. RNA in blood is altered prior to hemorrhagic transformation in ischemic stroke. *Ann Neurol* 2013;74(2):232–40.
  10. O'Connell GC, Treadway MB, Petrone AB, et al. Peripheral blood AKAP7 expression as an early marker for lymphocyte-mediated post-stroke blood brain barrier disruption. *Sci Rep* 2017;7(1):1172–1172.
  11. Whiteley W, Tseng M-C, Sandercock P. Blood biomarkers in the diagnosis of ischemic stroke: a systematic review. *Stroke J Cereb Circ* 2008;39(10):2902–9.
  12. Stafford P, Cichacz Z, Woodbury NW, Johnston SA. Immunosignature system for diagnosis of cancer. *Proc Natl Acad Sci* 2014;111(30):E3072–80.
  13. Hughes AK, Cichacz Z, Scheck A, Coons SW, Johnston SA, Stafford P. Immunosignaturing can detect products from molecular markers in brain cancer. *Marques ET, editor. PLoS ONE*. 2012;7(7):e40201.
  14. Legutki JB, Zhao Z-G, Greving M, Woodbury N, Johnston SA, Stafford P. Scalable high-density peptide arrays for comprehensive health monitoring. *Nat Commun* 2014;5:4785.
  15. Kidwell CS, Warach S. Acute ischemic cerebrovascular syndrome: diagnostic criteria. *Stroke* 2003;34(12):2995–8.
  16. Ross I, Robert G, Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comput Graph Stat* 1996;5(3):299–314.
  17. Singh S, Stafford P, Schlauch KA, et al. Humoral immunity profiling of subjects with myalgic encephalomyelitis using a random peptide microarray differentiates cases from controls with high specificity and sensitivity. *Mol Neurobiol* [Internet]. 2016 Dec 15 [cited 2018 Jan 16]; Available from: <http://link.springer.com/10.1007/s12035-016-0334-0>.
  18. Liaw A, Wiener M. Classification and regression by randomForest. *R News* 2002;2(3):18–22.
  19. Banerjee M, Ding Y, Noone A. Identifying representative trees from ensembles. *Stat Med* 2012;31(15):1601–16.
  20. Grömping U. Relative importance for linear regression in R: the package relaimpo. *J Stat Softw* 2006;17(1):139–47.
  21. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12(1):1–8.
  22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 1995;289–300.
  23. Dambinova SA, Bettermann K, Glynn T, et al. Diagnostic potential of the NMDA receptor peptide assay for acute ischemic stroke. *PLoS ONE* 2012;7(7):1–7.
  24. Foerch C. Serum glial fibrillary acidic protein as a biomarker for intracerebral haemorrhage in patients with acute stroke. *J Neurol Neurosurg Psychiatry* 2006;77(2):181–4.
  25. Reynolds MA, Kirchick HJ, Dahlen JR, et al. Early biomarkers of stroke. *Clin Chem* 2003;49(10):1733–1739.
  26. Purrucker JC, Hametner C, Engelbrecht A, Bruckner T, Popp E, Poli S. Comparison of stroke recognition and stroke severity scores for stroke detection in a single cohort. *J Neurol Neurosurg Psychiatry* 2015;86(9):1021–8.
  27. Studnek JR, Asimos A, Dodds J, Swanson D. Assessing the validity of the Cincinnati prehospital stroke scale and the medic prehospital assessment for code stroke in an Urban Emergency Medical Services Agency. *Prehosp Emerg Care* 2013;17(3):348–53.
  28. Planas AM, Gomez-Choco M, Urra X, Gorina R, Caballero M, Chamorro A. Brain-derived antigens in lymphoid tissue of patients with acute stroke. *J Immunol* 2012;188(5):2156–63.
  29. Bornstein NM, Aronovich B, Korczyn A D, Shavit S, Michaelson DM, Chapman J. Antibodies to brain antigens following stroke. *Neurology* 2001;56(4):529–30.
  30. Kaley-Zylinska ML, Symes W, Little KCE, et al. Stroke patients develop antibodies that react with components of N-methyl-D-aspartate receptor subunit 1 in proportion to lesion size. *Stroke* 2013;44(8):2212–9.
  31. Becker KJ, Tanzi P, Zierath D, Buckwalter MS. Antibodies to myelin basic protein are associated with cognitive decline after stroke. *J Neuroimmunol* 2016;295–296:9–11.
  32. Barclay WS, Al-Nakib W, Higgins PG, Tyrrell DAJ. The time course of the humoral immune response to rhinovirus infection. *Epidemiol Infect* 1989;103(3):659–669.
  33. Grau AJ, Urbanek C, Palm F. Common infections and the risk of stroke. *Nat Rev Neurol* 2010;6(12):681–94.
  34. Urbanus RT, Siegerink B, Roest M, Rosendaal FR, de Groot PG, Algra A. Antiphospholipid antibodies and risk of myocardial infarction and ischaemic stroke in young women in the RATIO study: a case-control study. *Lancet Neurol* 2009;8(11):998–1005.
  35. Wiseman SJ, Ralston SH, Wardlaw JM. Cerebrovascular disease in rheumatic diseases: a systematic review and meta-analysis. *Stroke*. 2016;47(4):943–950.
  36. Xiao Z, Pei Z, Yuan M, Li X, Chen S, Xu L. Risk of stroke in patients with inflammatory bowel disease: a systematic review and meta-analysis. *J Stroke Cerebrovasc Dis* 2015;24(12):2774–80.
  37. Deb P, Sharma S, Hassan KM. Pathophysiologic mechanisms of acute ischemic stroke: an overview with emphasis on therapeutic significance beyond thrombolysis. *Pathophysiology* 2010;17(3):197–218.
  38. Broger T, Basu Roy R, Filomena A, et al. Diagnostic performance of tuberculosis-specific igg antibody profiles in patients with presumptive tuberculosis from two continents. *Clin Infect Dis* 2017;64(7):947–55.
  39. Breiman L. Out-of-bag estimation [Internet]. 1996. Available from: <https://www.stat.berkeley.edu/~breiman/OOBestimation.pdf>.