Expression of Cytokines and Chemokines as Predictors of Stroke Outcomes in Acute Ischemic Stroke

Sarah R. Martha[1], Qiang Cheng[2], Justin F. Fraser[3,4,5,6,7], Liyu Gong[2], Lisa A. Collier[3], Stephanie M. Davis[3], Doug Lukins[4,5,6,7], Abdulnasser Alhajeri[3,7], Stephen Grupke[5,7] and Keith R. Pennypacker[3,6]*

[1] School of Nursing, University of Washington, Seattle, WA, United States, [2] Institute for Biomedical Informatics, University of Kentucky, Lexington, KY, United States, [3] Department of Neurology, University of Kentucky, Lexington, KY, United States, [4] College of Medicine, University of Kentucky, Lexington, KY, United States, [5] Departments of Neurosurgery, University of Kentucky, Lexington, KY, United States, [6] Neuroscience, University of Kentucky, Lexington, KY, United States, [7] Radiology, University of Kentucky, Lexington, KY, United States

Introduction: Ischemic stroke remains one of the most debilitating diseases and is the fifth leading cause of death in the US. The ability to predict stroke outcomes within the acute period of stroke would be essential for care planning and rehabilitation. The Blood and Clot Thrombectomy Registry and Collaboration (BACTRAC; clinicaltrials.gov NCT03153683) study collects arterial blood immediately distal and proximal to the intracranial thrombus at the time of mechanical thrombectomy. These blood samples are an innovative resource in evaluating acute gene expression changes at the time of ischemic stroke. The purpose of this study was to identify inflammatory genes and important immune factors during mechanical thrombectomy for emergent large vessel occlusion (ELVO) and which patient demographics were predictors for stroke outcomes (infarct and/or edema volume) in acute ischemic stroke patients.

Methods: The BACTRAC study is a non-probability sampling of male and female subjects ($\geq$18 year old) treated with mechanical thrombectomy for ELVO. We evaluated 28 subjects (66 ± 15.48 years) relative concentrations of mRNA for gene expression in 84 inflammatory molecules in arterial blood distal and proximal to the intracranial thrombus who underwent thrombectomy. We used the machine learning method, Random Forest to predict which inflammatory genes and patient demographics were important features for infarct and edema volumes. To validate the overlapping genes with outcomes, we perform ordinary least squares regression analysis.

Results: Machine learning analyses demonstrated that the genes and subject factors CCR4, IFNA2, IL-9, CXCL3, Age, T2DM, IL-7, CCL4, BMI, IL-5, CCR3, TNFα, and IL-27 predicted infarct volume. The genes and subject factor IFNA2, IL-5, CCL11, IL-17C, CCR4, IL-9, IL-7, CCR3, IL-27, T2DM, and CSF2 predicted edema volume. The overlap of genes CCR4, IFNA2, IL-9, IL-7, IL-5, CCR3, and IL-27 with T2DM predicted both infarct and edema volumes. These genes relate to a microenvironment for chemoattraction and proliferation of autoimmune cells, particularly Th2 cells and neutrophils.

Conclusions: Machine learning algorithms can be employed to develop prognostic predictive biomarkers for stroke outcomes in ischemic stroke patients, particularly in regard to identifying acute gene expression changes that occur during stroke.

Keywords: ischemic stroke, machine learning, gene expression, cytokines, chemokines

INTRODUCTION

Ischemic stroke is the fifth cause of death in the United States, accounting for 87% of all strokes (1). Approximately 800,000 individuals are affected per year and ischemic stroke is the primary cause of severe long-term disability (1). When a major cerebral artery is blocked by a thrombus, the occlusion causes deprivation of oxygen, glucose, and other essential nutrients to the brain which results in necrotic cerebral tissue (2). Current treatments focused on recanalization of the occluded cerebral artery for emergent large vessel occlusions (ELVO) via intravenous administration of tissue plasminogen activator (tPA) and/or endovascular mechanical thrombectomy (3–5). Stroke patients must be given tPA within a narrow window of 4.5 h, while mechanical thrombectomy has an extended window of intervention up to 24 h post-stroke (6–8). However, ischemic stroke patients must meet particular imaging or time criteria to receive either or both treatments, and many patients are out of the "window of opportunity," and receive only supportive medical care (9).

With early reperfusion to the cerebral artery, at risk tissue in the penumbra may be salvaged (10, 11). Even if blood flow is restored, neurons in the penumbra face major challenges to their survival, such as acidosis, ion pump failure, excitotoxicity, inflammatory response, and the breakdown the blood brain barrier (BBB) (12–14). Inflammation following an acute ischemic stroke is multifaceted, and is part of the normal pathological processes involved in the post-ischemic brain. Within the first few hours, there is a rapid activation of resident microglia that start to release proinflammatory mediators from the ischemic endothelium and brain parenchyma [including interleukin-(IL)1β, IL-6, IL-12, IL-23, and tumor necrosis factor-α (TNF-α)] (15–17). Furthermore, there is an activation of peripheral immune cells including neutrophils and T-cells that can cross the damaged BBB and accumulate at the injury site in the microvascular environment. In this acute phase of ischemic stroke more extensive damage occurs with the associated increased intracranial pressure leading to malignant cerebral edema, neuronal death, hemorrhagic transformation, and mortality (18–21).

Mechanical thrombectomy has greatly improved the ability to restore blood flow and improve patient outcomes, even when used in selected patients up to 24 h post-stroke onset (5, 22). The process of thrombectomy permits us to extract blood directly in the vicinity of the infarct from a stroke patient to provide the first glimpse into cellular/molecular events occurring during an ischemic stroke. Through the standard thrombectomy process, we can isolate distal blood within the artery immediately downstream from the clot, peripheral blood that is proximal to the clot (systemic arterial blood in the cervical carotid artery), and the thrombus itself upon removal from the human subject (23).

Understanding of molecular and cellular changes intraluminally in ischemic stroke may provide develop deeper insights into the inflammatory pathways involved, and could lead to the identification of prognostic biomarkers in stroke patients. This information could potentially provide benchmarks to study responses to treatment and aid in development of new therapies. The purpose of this study was to identify inflammatory genes and important immune mediators at the time of mechanical thrombectomy and which patient demographics were predictors for stroke outcomes (infarct and/or edema volume) in acute ischemic stroke patients.

METHODS

Study Design

This study was developed from the Blood and Clot Thrombectomy Registry and Collaboration (BACTRAC) protocol: a prospectively enrolling tissue bank for emergent large vessel occlusion (ELVO) (clinicaltrials.gov NCT03153683) (23). The BACTRAC study is a non-probability sampling of subjects over 18 years old, conducted at a large academic comprehensive stroke center in central Kentucky, and who were treated with mechanical thrombectomy for ELVO ischemic stroke. While the BACTRAC tissue bank is continuously enrolling, data for this study was collected from subjects enrolled May 2017 through January 2019. The Institutional Review Board approved this study, and informed consent was obtained from all subjects or legally authorized representatives.

Sample Population

Subjects ($\geq$18 years old) were eligible for inclusion if they were hospitalized with a diagnosis for an ischemic stroke and underwent thrombectomy procedure for ELVO. Subjects were candidates for inclusion if they met criteria: informed consent was signed and dated by subject or legally authorized representative within 24 h post-thrombectomy, confirmation of acute ischemic stroke based on clinical and radiographic imaging (CT and/or MRI), intra-arterial thrombectomy as determined by neurointerventional radiology (NIR) physician, an acute thromboembolus within an intracranial artery and underwent endovascular thrombectomy procedure. Subjects were excluded if baseline urine pregnancy test was positive and/or breastfeeding and/or concurrently participated in treatment studies.

Data Collection

A medical record review was conducted to collect admission clinical and demographic variables. These

variables included age, sex, ethnicity, height, weight, body mass index (BMI), comorbidities, smoking status, National Institutes of Health Stroke Scale (NIHSS), baseline computerized tomography (CT) and/or magnetic resonance imaging (MRI) scan to verify large vessel occlusion stroke/thrombectomy eligibility and CTA collateral score, and time of last known normal (LKN) symptom (prior to ischemic stroke).

Tissue removal and tissue banking protocols were previously published in detail (23). In brief, this was a collaborative effort among the NeuroInterventional Radiology (NIR) service line and our Center for Advanced Translational Stroke Science. At the beginning of the mechanical thrombectomy procedure, during initial access through the clot to the distal intracranial vessels, the NIR physician backbleeds the microcatheter to determine if they are distal to the distal end of the thrombus. Through this, 1 ml of whole arterial blood from the microcatheter (0.021 inch inner lumen diameter) was collected. Simultaneously, the collection of 7 ml of whole arterial peripheral blood was collected from the cervical parent artery (internal carotid artery or vertebral artery) during backbleeding while initially suction for thrombectomy. The distal (intracranial) and proximal (peripheral) blood samples were handed off to the on-call research associates who immediately processed the specimens in the wet laboratory adjacent to the angiography suite.

An additional medical chart review was conducted to gather post-thrombectomy endpoint variables: last known normal (LKN) symptom to thrombectomy completion time (infarct time), thrombolysis in cerebral infarction (TICI) scores, and NIHSS at discharge. A blinded neuroradiologist evaluated post-thrombectomy computed tomography (CT) imaging and/or magnetic resonance imaging (MRI) for infarct and edema volumes, and hemorrhagic grade. Clinical and demographic variables, along with post-thrombectomy endpoints were deidentified and maintained in our institution's REDCap software.

RNA Extraction and Amplification

The extraction of total RNA from the cellular pellet/buffy coat was collected from the stored distal and proximal arterial blood samples using the Nucleospin Blood Kit (Macherey-Nagel, Düren, Germany), per manufacturer protocol. Nanodrop Lite Spectrophotometer (Thermo-Fisher Scientific; Waltham, MA) estimated the quantity of total RNA based on the A260/A280 ratio. To validate total RNA, RNA integrity numbers (RIN) values were obtained from the hospital's Genomics Core facility. The assessment and analysis of RNA integrity was evaluated with Eukaryote Total RNA Nano assay on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and with associated Agilent 2100 Bioanalyzer expert software. The mean RIN values (±SD) for distal and proximal arterial blood samples were 8.89 ± 0.41 (range from 8.30 to 9.60) and 6.98 ± 0.29 (range from 6.60 to 7.50), respectfully. The RIN values indicate moderate to high quality RNA with minimal degradation (24). RT^2 First Strand Kit (Qiagen) was used to synthesize cDNA which were analyzed for gene expression using Qiagen RT^2 Profiler PCR Array Human Inflammatory Cytokines & Receptors

for 84 genes associated with inflammation and evaluated with StepOnePlus Real-Time PCR System interfaced with StepOne software (Applied Biosystems, Foster City, CA). Each gene array plate had five housekeeping genes (ACTB, B2M, GAPDH, HPRT1, and RPLP0) to normalize the gene expression data by using $\Delta\Delta CT$ method. The distal blood data for each gene was compared to proximal blood data for each subject. Proximal blood samples were used as an internal control for each patient, since there is not a true control. Since the cervical segments have circulating blood, we believe these are equivalent to arterial blood taken from peripheral arterial location.

Data Analysis

Descriptive and frequency statistics were obtained and used to characterize male and female ischemic stroke subjects. Comparisons were made for subject demographics and baseline characteristics using SPSS, version 26 software (IBM, Armonk, NY). Statistical significance (p-value of 0.05) for group differences were assessed by independent t-tests for continuous variables and Chi-square or Fisher exact tests of associations based on categorical (ordinal and binary) variables. The gene expression mean fold change values were assessed with Scikit-learn Python package (v0.21.2) for machine learning analysis (25). Statesmodels Python package (v0.10.1) was used for ordinary least squares (OLS) regression analysis (26).

Machine Learning

For machine learning analysis, the demographics data used included variables: sex, age, race/ethnicity, BMI, comorbidities (hypertension, type 2 diabetes, hyperlipidemia, previous stroke, and atrial fibrillation), and two stroke outcomes (infarct and edema volumes). The gene expression dataset included 84 inflammatory genes. Any variables in the two stroke outcome datasets were dropped if the proportion of missing values were more than 40%. After pre-processing 89 variables were in the final dataset.

Predicting Infarct Volume

Predictive analysis was performed by using the infarct volume as the response variable, and all gene expression variables as predictors. The analysis was adjusted by using demographic variables as covariates. Regression was used to perform the predictive analysis task. Since there was a relatively small sample size with a mild number of predictors (i.e., 28 subjects with 89 predictors), a widely used machine learning method, Random Forest (RF) (27, 28) was employed to complete the analysis. This method was used because of the reliable performance on small sample datasets with mixed categorical and continuous variables.

RF is an ensemble of decision trees, where each internal node represents a condition, or test, on a single predictor, aiming to split the dataset into two so that each had a similar response values. RF provides versatile, accurate, and stable predictions for supervised machine learning tasks such as regression as well as measuring the relative importance of each predictor. The importance score of a predictor is calculated by how much the variance is reduced or information gain is increased across all

nodes that were used. The importance scores were scaled so the sum equals one.

Hyper-parameters of RF were selected in a coarse-to-fine approach using *k*-fold cross-validation (CV), which partitions the training dataset into *k* parts evenly, then rotationally uses *k-1* parts to train the machine learning model and tests the model with the remaining one part. The parameters were first selected using the 5-fold CV, then tuned the parameters around the selected value using the 10-fold CV. Predicting performance of RF was optimized in mean squared error (MSE), which is minimized for the optimal model.

With the selected hyper-parameters, the importance of features using RF were ranked. Ranking may be affected by randomness of RF due to the correlations between some features. Therefore, we programed RF 100 times and collected the mean values of the feature importance. The number of possible orders of the features is combinatorically large, this approach can effectively reduce but cannot fully eliminate the effect of the correlation between features.

Predicting Edema Volume

Similar to the prediction of infarct volumes, this was a regression task by using edema volumes as a response variable and the remaining gene variables as predictors. The analysis was adjusted using the demographic variables as covariates as done for the analysis of infarct volumes. Using similar preprocessing and cross validation with RF, the negative MSE was used as a score and maximized for optimal RF model. The relative importance of the predictors were also obtained for this task.

Stroke Outcomes of Infarct and Edema Volumes

Noncontrast head CT and CTA of the head and neck were obtained upon presentation to the emergency department during initial assessment for acute ischemic stroke. Siemens SOMATOM Definition Edge and SOMATOM Force CT scanners were used for all CT studies. CTA collateral scores were determined using maximum intensity projection images from CTA of the head with a scoring system described in previous work by Souza et al. (29). MRI and CT of the head without contrast were obtained following thrombectomy. MRI was performed using Siemens MAGNETOM Aera and MAGNETOM Skyra machines at magnetic field strength of 1.5 and 3.0 Tesla, respectively. Hemorrhage grade, infarct volume, and edema volume were determined on post-thrombectomy MRI of the head or CT of the head if MRI was unavailable. Hemorrhage grade was determined using a grading scale described by Hacke et al. (30). Infarct volumes and edema volumes were calculated using post-thrombectomy MRI, or using CT if MRI was unavailable. Imaging included the entire brain on both MRI and CT examinations, and all images/slices were visually assessed for infarction or edema. When MRI was available, diffusion weighted images (DWI) were used to calculate infarct volumes, and T2 FLAIR images were used to calculate edema volumes. The areas of abnormal signal (restricted diffusion on DWI or hyperintense signal on T2 FLAIR) were manually segmented and analyzed to determine volume using ITK-SNAP

software (www.itksnap.org) (31). When MRI was unavailable, CT of the head was used to calculate both infarct and edema volume. As infarct and edema were indistinguishable on CT, the same value was used for both after manually segmenting the area of abnormal hypoattenuation using ITK-SNAP. All imaging assessment, including hemorrhage grade, CTA collateral score, infarct volume, and edema volume was performed by an experienced neuroradiologist in a blinded fashion.

RESULTS

Sample Characteristics

Twenty-eight subjects underwent mechanical thrombectomy during the study period. The mean age of our subjects were 66 ± 15.48 years, primarily female (61%), and the majority were Caucasian (86%). **Table 1** presents baseline subject characteristics and comparisons based on sex. Similarities were exhibited between the subjects for BMI, comorbidities (hypertension, atrial fibrillation, type 2 diabetes, hyperlipidemia, previous stroke, chronic obstructive pulmonary disease, and coronary artery disease), smoking status, stroke location, hemorrhagic grade, CTA collateral scores, infarct and edema volumes. The NIHSS on average went from admission score of 17 ± 6.56 to discharge score of 11 ± 9.54 post-thrombectomy. More males (81.8%) than females (41.2%) exhibited full perfusion to the occluded artery post-thrombectomy with a TICI score of 3 ($p = 0.034$). The mean infarct time (LKN to thrombectomy completion time) for males was ~7 h vs. almost 10 h for females ($p = 0.093$).

Prediction of Infarct Volume With Machine Learning

For predicting infarct volume, the covariates including all 84 genes and the 10 demographic variables are all ranked in importance scores. For the ranked top *k, k = 1–84,* covariates we perform 5-fold CVs to find the optimal RF model and obtain the corresponding optimal performance, which is scored in negative MSE. The importance values of the top 10 covariates, 10 for predictive analysis of infarct volume are given in **Table 2**. They are in descending order as follows: CCR4, IFNA2, IL-9, CXCL3, age, type 2 diabetes (T2DM), IL-7, CCL4, BMI, IL-5, CCR3, TNFα, and IL-27.

Prediction of Edema Volume With Machine Learning

For predictive analysis of edema volume, in a similar way to the predictive analysis of infarct volume, we obtain the optimal performance for the top *k* covariates, *k* = 1–84. The importance values of the top 10 most important covariates are shown in **Table 3**. They are in descending order as follows: IFNA2, IL-5, CCL11, IL-17C, CCR4, IL-9, IL-7, CCR3, IL-27, T2DM, and CSF2.

When using infarct and edema volumes as outcome variables, we observed selected gene subsets for these two predictive analysis have a significant overlap. The intersection of the top 10 genes from the two stroke outcomes are seven genes with T2DM, these include CCR4, IFNA2, IL-9, IL-7, IL-5, CCR3, and

TABLE 1 | Baseline demographics and characteristics for ischemic stroke subjects.

	Males (n = 11)	Females (n = 17)	p-value
Age (years)	68 ± 17.07	64 ± 14.77	0.606
Ethnicity			
African-American	2 (18.2)	0 (0.0)	0.072
Caucasian	9 (81.8)	15 (88.2)	0.650
Unknown	0 (0.0)	2 (11.8)	0.254
BMI			
Under/normal weight	4 (36.4)	9 (52.9)	0.409
Overweight	6 (54.5)	6 (35.3)	0.333
Obese	0 (0.0)	0 (0.0)	–
Morbidly obese	1 (9.1)	2 (11.8)	0.831
Comorbidities			
Hypertension	4 (36.4)	11 (64.7)	0.346
Atrial fibrillation	3 (27.3)	5 (29.4)	0.907
Diabetes	1 (9.1)	6 (35.3)	0.127
Hyperlipidemia	3 (27.3)	2 (11.8)	0.313
Previous stroke	2 (18.2)	3 (17.6)	0.973
COPD	1 (9.1)	3 (17.6)	0.545
CAD	3 (27.3)	1 (5.9)	0.123
Smoking status			
Never	5 (45.5)	11 (64.7)	0.333
Currently	4 (36.4)	4 (23.5)	0.481
Previously (>6 months)	2 (18.2)	2 (11.8)	0.650
NIHSS on admission	17 ± 8.04	16 ± 5.65	0.732
Minor stroke (1-4)	2 (18.2)	1 (5.9)	0.322
Moderate stroke (5-15)	3 (27.3)	6 (35.3)	0.671
Moderate/severe (16-20)	2 (18.2)	6 (35.3)	0.346
Severe stroke (≥21)	4 (36.4)	4 (23.5)	0.481
NIHSS at discharge	14 ± 14.21	9 ± 7.32	0.249
Minor stroke (1-4)	4 (36.4)	6 (35.3)	0.808
Moderate stroke (5-15)	4 (36.4)	7 (41.2)	0.818
Moderate/severe (16-20)	1 (9.1)	2 (11.8)	0.838
Severe stroke (≥21)	2 (18.2)	1 (5.9)	0.322
TICI score			
2A ≤ 50% Perfusion	0 (0.0)	1 (5.9)	0.432
2B ≥ 50% Perfusion	2 (18.2)	9 (52.9)	0.070
3 = Full perfusion	9 (81.8)	7 (41.2)	0.034
LKN to thrombectomy completion time (minutes)	416 ± 231.88	582 ± 249.06	0.093
Stroke location			
Left MCA	4 (36.4)	8 (47.1)	0.576
Right MCA	6 (54.5)	8 (47.1)	0.699
Basilar	1 (9.1)	1 (5.9)	0.747
Infarct volume (cm³)	87.90 ± 104.58	48.33 ± 62.30	0.219
Edema volume (cm³)	98.60 ± 116.01	48.52 ± 61.23	0.147
Hemorrhagic grade None	4 (36.4)	4 (23.5)	0.481
HI1	4 (36.4)	7 (41.2)	0.808
HI2	1 (9.1)	6 (35.3)	0.127
PH1	1 (9.1)	0 (0.0)	0.220
PH2	1 (9.1)	0 (0.0)	0.220

(Continued)

TABLE 1 | Continued

	Males (n = 11)	Females (n = 17)	p-value
CTA collateral score			
0	3 (27.3)	2 (11.8)	0.650
1	5 (45.5)	11 (64.7)	0.333
2	3 (27.3)	4 (23.5)	0.748

Values are mean ± SD or (%). Comparisons were performed with independent t-tests, Chi-square, or Fisher exact tests based on distribution of data.

TABLE 2 | Importance values of top 10 variables for predicting infarct volume.

CCR4	0.077
IFNA2	0.063
IL-9	0.053
CXCL3	0.045
Age	0.038
Type 2 diabetes	0.036
IL-7	0.034
CCL4	0.029
BMI	0.028
IL-5	0.027
CCR3	0.023
TNFα	0.022
IL-27	0.021

TABLE 3 | Importance values of top 10 variables for predicting edema volume.

IFNA2	0.074
IL-5	0.070
CCL11	0.057
IL-17C	0.045
CCR4	0.045
IL9	0.044
IL-7	0.036
CCR3	0.035
IL-27	0.033
Type 2 diabetes	0.032
CSF2	0.025

IL-27. This overlapping subset of genes were regarded as the most important associations at the time of mechanical thrombectomy.

Prediction of Stroke Outcomes With Multiple Linear Regression

To validate this overlapping subset of genes, we perform ordinary least squares (OLS) regression analysis with the top seven covariates. The OLS results for predictors of infarct volume and edema volume are shown in **Tables 4**, **5**, respectively. To note, the OLS was a statistical model that assumes a linear relationship between the predictive outcome and the covariates. In contrast,

TABLE 4 | OLS regression variables predicting infarct volume ($n = 28$).

Variable	β	P-value
CCR4	−0.733	0.740
IFNa2	−0.104	0.793
IL-9	12.146	0.017
Type 2 diabetes	69.942	0.026
IL-7	−5.277	0.217
IL-5	1.205	0.233
CCR3	−0.084	0.855
IL-27	−0.683	0.549

$R^2 = 0.654$, adjusted $R^2 = 0.533$, df = 8, F = 5.396, p = 0.001.
Durbin-Watson = 1.83.

TABLE 5 | OLS regression variables predicting edema volume ($n = 28$).

Variable	β	P-value
CCR4	−1.987	0.486
IFNa2	−0.099	0.848
IL-9	13.127	0.040
Type 2 Diabetes	57.469	0.140
IL-7	−8.326	0.134
IL-5	2.109	0.110
CCR3	0.083	0.889
IL-27	−0.729	0.693

$R^2 = 0.507$, adjusted $R^2 = 0.334$, df = 8, F = 2.938, p = 0.028.
Durbin-Watson = 1.69.

the machine learning model was capable of capturing potentially highly non-linear relationship between the outcome and the covariates (32, 33). Therefore, OLS may partially capture the association between the outcomes and the covariates. IL-9 and T2DM were predictors for infarct volume [F(8,27 = 5.396, $p = 0.001$). These predictors for infarct volume explained 53% of the total variance in the model (**Table 4**). Similarly, we performed the OLS analysis for predictors of edema volume. IL-9 was a predictor for edema volume [F(8,27 = 2.938, $p = 0.028$). IL-9 explains 33% of the total variance found in the model (**Table 5**).

Gene Expression of Inflammatory Markers and Ingenuity Pathways Analysis

Ten inflammatory genes (CCR4, IFNA2, IL-9, CXCL3, IL-7, CCL4, IL-5, CCR3, TNFα, and IL-27) were recognized as predictors for infarct volume had mean fold change ranging from 2.56 to 35.06. While the inflammatory genes (IFNA2, IL-5, CCL11, IL-17C, CCR4, IL-9, IL-7, CCR3, IL-27, and CSF2) that were identified as predictors for edema volume had mean fold change ranging from 3.98 to 42.61. The seven important predictor genes (CCR4, IFNA2, IL-9, IL-7, IL-5, CCR3, and IL-27) that overlapped both stroke outcomes had mean fold change ranging from 3.98 to 35.06.

The mean fold changes for the 13 genes were analyzed through the use of Ingenuity Pathway Analysis (IPA) software (Qiagen Inc., https://www.qiagenbioinformatics.com/products/
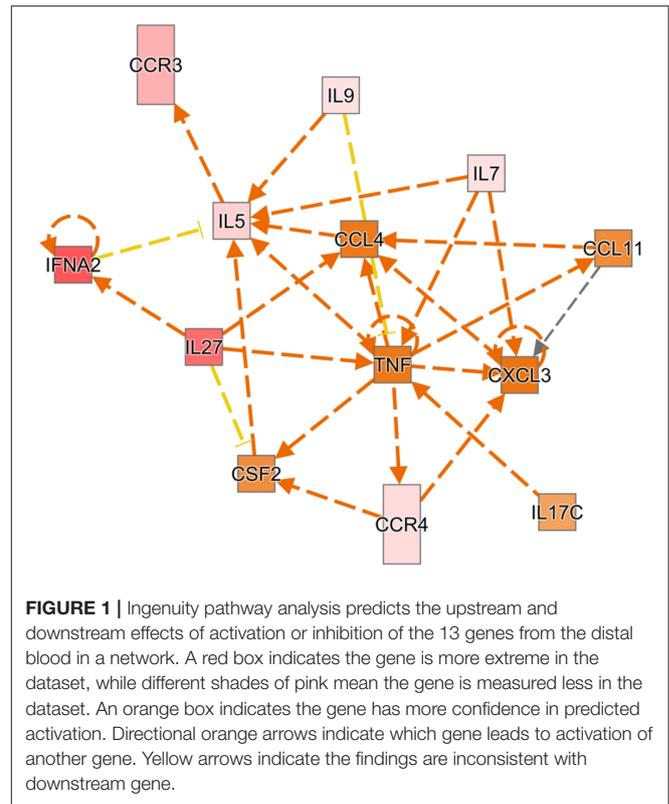
FIGURE 1 | Ingenuity pathway analysis predicts the upstream and downstream effects of activation or inhibition of the 13 genes from the distal blood in a network. A red box indicates the gene is more extreme in the dataset, while different shades of pink mean the gene is measured less in the dataset. An orange box indicates the gene has more confidence in predicted activation. Directional orange arrows indicate which gene leads to activation of another gene. Yellow arrows indicate the findings are inconsistent with downstream gene.

ingenuity-pathway-analysis) (34). IPA predicts the upstream and downstream effects of activation or inhibition of the 13 genes from the distal blood in a pathway network. The overall inflammatory chemokine/cytokine pattern from the distal blood characterizes a microenvironment for chemoattraction and proliferation of immune cells, particularly in Th2 response and neutrophils (**Figure 1**).

DISCUSSION

Our results revealed insights to the initial inflammatory response to ischemic stroke in stroke patients. Arterial blood samples were collected at one of the earliest timepoints (∼8 h and 35 min from LKN to thrombectomy completion time or infarct time) during an ongoing stroke. Expression of 13 cytokines and chemokines were induced over 2-fold in the distal blood relative to levels in the proximal blood. Machine learning identified 10 genes, along with age, T2DM, and BMI for predictors of infarct volume and 10 genes with T2DM for predictors of edema volume. In addition, there was an overlap with seven genes and T2DM as predictors for both stroke outcomes.

These seven genes are related to the Th2 response, which is a T-cell response associated with allergy, cancer, and autoimmunity (35, 36). One study has shown a systemic shift of the immune system toward Th2 response at the late post-acute phase in stroke patients after examining their peripheral venous blood (37). IFNA2 is primarily produced by dendritic cells (38). This cytokine is targeted as a treatment for the autoimmune disease,

systemic lupus erythematosus (39). Exposure to INFA2 has been linked to causing edema and been shown to induce the production of cytokine IL-7 (40–42). IL-7 is required for T-cell development and function (43, 44). The expression of IL-4, a classic marker for Th2 cells, is increased by IL-7 and is a direct cause of edema (45, 46). This cytokine's activity is targeted for diseases related to pathogenic T-cells (47). Th2 cells express both CCR3 and CCR4 receptors that bind chemokines to direct their migration to the site of injury or infection (48, 49). CCR3 is a chemokine receptor that has been shown to mediate neuronal injury after ischemic insult (50). At least eight different chemokines recognize this receptor and has shown that its activation enhances choroidal neovascularization and trafficking of eosinophils (51). Pharmaceutical companies have targeted this receptor for the treatment of severe asthma (52). CCR4 is another chemokine receptor and CCL17 and CCL22 are identified as high affinity ligands (53). This receptor is highly expressed in thymus and peripheral blood leukocytes, and is targeted for treatment in T-cell leukemia and peripheral T-cell lymphoma (54). The cytokines, IL-5 and IL-9 are both expressed and secreted by Th2 cells and are linked to edema and infarct volume, respectively (55, 56). IL-5 is produced in female mice express higher levels in the spleen relative to males (57). Antagonists to IL-5 have been produced for the treatment of severe asthma (58). Patients suffering from asthma have higher levels of IL-9 producing T-cells in their peripheral blood (59). While not targeted in a current treatment, systemically applied neutralizing antibodies reduces infarct (56). Cytokine, IL-27 has both pro-and anti-inflammatory effects and the main source of activation is by antigen-presenting cells (60). IL-27 activates an inflammatory response by inducing $CD4^+$ T-cells into a Th1 response (61–64). However, IL-27 can also suppress Th17 differentiation and inhibit inflammation enhancing a Th2 response (65–68). In an experimental stroke model, mice were administered IL-27 injections post-intracranial hemorrhage (ICH) and demonstrated improved behavior tests, mobility, and reduced edema around the hemorrhage sites (69).

Several limitations to this study should be noted. The sample for this study was small and subjects were mostly Caucasian due to hospital location. There was a potential for sampling bias since this was a non-probability strategy and based on completed mechanical thrombectomy procedures. A more diverse and larger sample size is warranted in future studies. Standard of care imaging for infarct and edema volumes were based on CT and/or MRI imaging post-thrombectomy. Additional imaging at discharge for comparison is warranted to understand how thrombectomy relates to these stroke outcomes. Follow-up assessment would be helpful to address subject's long-term recovery and functional status.

CONCLUSION

The machine learning (ML) algorithm has uncovered a molecular/cellular signaling pathway that the literature supports its relationship with the facilitation of infarct volume and edema. Despite significant funding for stroke drug discovery, tPA and mechanical thrombectomy are the only treatments. Identifying the initial activators of the inflammatory response to stroke could be therapeutic targets for treatment. The inhibition of these early inflammatory events has the potential to block the inflammatory cascade that occurs after a stroke that recruits the peripheral immune system, which is involved in neurodegeneration that lasts days afterward. The ML method has determined several related genes are predictors of edema and infarct volume in acute ischemic stroke patients. Thus, the use of ML to determine molecular predictors from ischemic blood could initiate a new approach for drug discovery in the field of stroke. Future studies will utilize ML methods and examine expression of genes that predict long-term functional outcomes.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by UK HealthCare Office of Research Integrity IRB # 17-0019-F1V clinicaltrials.gov NCT03153683. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SM, JF, and KP: conceptualization, review, and editing. SM, LC, SD, QC, LG, DL, AA, SG, JF, and KP: methodology. QC and LG: formal analysis. SM, QC, LG, DL, and KP: written original draft. All authors made a substantial and intellectual contribution to the work.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics-2019 update: A Report From the American Heart Association. *Circulation.* (2019) 139:e56–528. doi: 10.1161/CIR.0000000000000659

2. Borgens RB, Liu-Snyder P. Understanding secondary injury. *Q Rev Biol.* (2012) 87:89–127. doi: 10.1086/665457

3. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med.* (1995) 333:1581–7. doi: 10.1056/NEJM199512143332401

4. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet.* (2016) 387:1723–31. doi: 10.1016/S0140-6736(16)00163-X

5. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med.* (2015) 372:11–20. doi: 10.1056/NEJMoa1411587

6. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K. 2018 Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* (2018) 49:e46–110. doi: 10.1161/STR.0000000000000158

7. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med.* (2018) 378:11–21. doi: 10.1056/NEJMoa1706442

8. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *N Engl J Med.* (2018) 378:708–18. doi: 10.1056/NEJMoa1713973

9. Fonarow GC, Zhao X, Smith EE, Saver JL, Reeves MJ, Bhatt DL, et al. Door-to-needle times for tissue plasminogen activator administration and clinical outcomes in acute ischemic stroke before and after a quality improvement initiative. *JAMA.* (2014) 311:1632–40. doi: 10.1001/jama.2014.3203

10. Doyle KP, Simon RP, Stenzel-Poore MP. Mechanisms of ischemic brain damage. *Neuropharmacology.* (2008) 55:310–8. doi: 10.1016/j.neuropharm.2008.01.005

11. Tobin MK, Bonds JA, Minshall RD, Pelligrino DA, Testai FD, Lazarov O. Neurogenesis and inflammation after ischemic stroke: what is known and where we go from here. *J Cereb Blood Flow Metab.* (2014) 34:1573–84. doi: 10.1038/jcbfm.2014.130

12. Heo JH, Lucero J, Abumiya T, Koziol JA, Copeland BR, del Zoppo GJ. Matrix metalloproteinases increase very early during experimental focal cerebral ischemia. *J Cereb Blood Flow Metab.* (1999) 19:624–33. doi: 10.1097/00004647-199906000-00005

13. Mifsud G, Zammit C, Muscat R, Di Giovanni G, Valentino M. Oligodendrocyte pathophysiology and treatment strategies in cerebral ischemia. *CNS Neurosci Ther.* (2014) 20:603–12. doi: 10.1111/cns.12263

14. Kristian T, Siesjo BK. Changes in ionic fluxes during cerebral ischaemia. *Int Rev Neurobiol.* (1997) 40:27–45. doi: 10.1016/S0074-7742(08)60714-X

15. Strbian D, Karjalainen-Lindsberg ML, Tatlisumak T, Lindsberg PJ. Cerebral mast cells regulate early ischemic brain swelling and neutrophil accumulation. *J Cereb Blood Flow Metab.* (2006) 26:605–12. doi: 10.1038/sj.jcbfm.9600228

16. Pradillo JM, Denes A, Greenhalgh AD, Boutin H, Drake C, McColl BW, et al. Delayed administration of interleukin-1 receptor antagonist reduces ischemic brain damage and inflammation in comorbid rats. *J Cereb Blood Flow Metab.* (2012) 32:1810–9. doi: 10.1038/jcbfm.2012.101

17. Savage CD, Lopez-Castejon G, Denes A, Brough D. NLRP3-inflammasome activating DAMPs stimulate an inflammatory response in glia in the absence of priming which contributes to brain inflammation after injury. *Front Immunol.* (2012) 3:288. doi: 10.3389/fimmu.2012.00288

18. Mokri B. The Monro-Kellie hypothesis: applications in CSF volume depletion. *Neurology.* (2001) 56:1746–8. doi: 10.1212/WNL.56.12.1746

19. Simard JM, Kent TA, Chen M, Tarasov KV, Gerzanich V. Brain oedema in focal ischaemia: molecular pathophysiology and theoretical implications. *Lancet Neurol.* (2007) 6:258–68. doi: 10.1016/S1474-4422(07)70055-8

20. Amantea D, Nappi G, Bernardi G, Bagetta G, Corasaniti MT. Post-ischemic brain damage: pathophysiology and role of inflammatory mediators. *FEBS J.* (2009) 276:13–26. doi: 10.1111/j.1742-4658.2008.06766.x

21. Dostovic Z, Dostovic E, Smajlovic D, Ibrahimagic OC, Avdic L. Brain edema after ischaemic stroke. *Med Arch.* (2016) 70:339–41. doi: 10.5455/medarh.2016.70.339-341

22. Albers GW, Lansberg MG, Kemp S, Tsai JP, Lavori P, Christensen S, et al. A multicenter randomized controlled trial of endovascular therapy following imaging evaluation for ischemic stroke (DEFUSE 3). *Int J Stroke.* (2017) 12:896–905. doi: 10.1177/1747493017701147

23. Fraser JF, Collier LA, Gorman AA, Martha SR, Salmeron KE, Trout AL. The Blood And Clot Thrombectomy Registry And Collaboration (BACTRAC) protocol: novel method for evaluating human stroke. *J Neurointerv Surg.* (2019) 11:265–70. doi: 10.1136/neurintsurg-2018-014118

24. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol.* (2006) 7:3. doi: 10.1186/1471-2199-7-3

25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* (2011) 12:2825–30.

26. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. In: *Proceedings of the 9th Python in Science Conference*, 2010 (2010).

27. Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324

28. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning.* New York, NY: Springer New York Inc. (2001).

29. Souza LC, Yoo AJ, Chaudhry ZA, Payabvash S, Kemmling A, Schaefer PW, et al. Malignant CTA collateral profile is highly specific for large admission DWI infarct core and poor outcome in acute stroke. *AJNR Am J Neuroradiol.* (2012) 33:1331–6. doi: 10.3174/ajnr.A2985

30. Hacke W, Kaste M, Fieschi C, von Kummer R, Davalos A, Meier D, et al. Randomised double-blind placebo-controlled trial of thrombolytic therapy with intravenous alteplase in acute ischaemic stroke (ECASS II). Second European-Australasian Acute Stroke Study Investigators. *Lancet.* (1998) 352:1245–51. doi: 10.1016/S0140-6736(98)08020-9

31. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage.* (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015

32. Vapnik VN. *The Nature of Statistical Learning Theory.* Newyork, NY: Springer-Verlag New York, Inc. (1995).

33. Vapnik VN. *The Nature of Statistical Learning Theory.* 2nd ed. Springer (2000).

34. Kramer A, Green J, Pollard J Jr., Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics.* (2014) 30:523–30. doi: 10.1093/bioinformatics/btt703

35. Christoffersson G, von Herrath M. Regulatory immune mechanisms beyond regulatory T cells. *Trends Immunol.* (2019) 40:482–91. doi: 10.1016/j.it.2019.04.005

36. Chraa D, Naim A, Olive D, Badou A. T lymphocyte subsets in cancer immunity: friends or foes. *J Leukoc Biol.* (2019) 105:243–55. doi: 10.1002/JLB.MR0318-097R

37. Theodorou GL, Marousi S, Ellul J, Mougiou A, Theodori E, Mouzaki A, et al. T helper 1 (Th1)/Th2 cytokine expression shift of peripheral blood CD4+ and CD8+ T cells in patients at the post-acute phase of stroke. *Clin Exp Immunol.* (2008) 152:456–63. doi: 10.1111/j.1365-2249.2008.03650.x

38. Hagberg N, Berggren O, Leonard D, Weber G, Bryceson YT, Alm GV, et al. IFN-alpha production by plasmacytoid dendritic cells stimulated with RNA-containing immune complexes is promoted by NK cells via MIP-1beta and LFA-1. *J Immunol.* (2011) 186:5085–94. doi: 10.4049/jimmunol.1003349

39. Vukelic M, Li Y, Kyttaris VC. Novel treatments in lupus. *Front Immunol.* (2018) 9:2658. doi: 10.3389/fimmu.2018.02658

40. Dusheiko G. Side effects of alpha interferon in chronic hepatitis C. *Hepatology.* (1997) 26:112s−21s. doi: 10.1002/hep.510260720

41. Rostaing L, Izopet J, Baron E, Duffaut M, Puel J, Durand D. Treatment of chronic hepatitis C with recombinant interferon alpha in kidney transplant recipients. *Transplantation.* (1995) 59:1426–31. doi: 10.1097/00007890-199505270-00012

42. Parker R, Dutrieux J, Beq S, Lemercier B, Rozlan S, Fabre-Mersseman V, et al. Interleukin-7 treatment counteracts IFN-alpha therapy-induced lymphopenia and stimulates SIV-specific cytotoxic T lymphocyte responses in SIV-infected rhesus macaques. *Blood.* (2010) 116:5589–99. doi: 10.1182/blood-2010-03-276261

43. ElKassar N, Gress RE. An overview of IL-7 biology and its use in immunotherapy. *J Immunotoxicol.* (2010) 7:1–7. doi: 10.3109/154769 10903453296

44. Moors M, Vudattu NK, Abel J, Kramer U, Rane L, Ulfig N, et al. Interleukin-7 (IL-7) and IL-7 splice variants affect differentiation of human neural progenitor cells. *Genes Immun.* (2010) 11:11–20. doi: 10.1038/gene.2009.77

45. Hameg A, Gouarin C, Gombert JM, Hong S, van Kaer L, Bach JF, et al. IL-7 up-regulates IL-4 production by splenic NK1.1+ and NK1.1- MHC class I-like/CD1-dependent CD4+ T cells. *J Immunol.* (1999) 162:7067–74.

46. Wang XM, Zhang YG, Li AL, Long ZH, Wang D, Li XX, et al. Expressions of serum inflammatory cytokines and their relationship with cerebral edema in patients with acute basal ganglia hemorrhage. *Eur Rev Med Pharmacol Sci.* (2016) 20:2868–71.

47. Ellis J, van Maurik A, Fortunato L, Gisbert S, Chen K, Schwartz A, et al. Anti-IL-7 receptor alpha monoclonal antibody (GSK2618960) in healthy subjects - a randomized, double-blind, placebo-controlled study. *Br J Clin Pharmacol.* (2019) 85:304–15. doi: 10.1111/bcp.13748

48. Xia MQ, Qin SX, Wu LJ, Mackay CR, Hyman BT. Immunohistochemical study of the beta-chemokine receptors CCR3 and CCR5 and their ligands in normal and Alzheimer's disease brains. *Am J Pathol.* (1998) 153:31–7. doi: 10.1016/S0002-9440(10)65542-3

49. Yamamoto S, Matsuo K, Nagakubo D, Higashiyama S, Nishiwaki K, Oiso N, et al. A CCR4 antagonist enhances DC activation and homing to the regional lymph node and shows potent vaccine adjuvant activity through the inhibition of regulatory T-cell recruitment. *J Pharmacol Sci.* (2018) 136:165–71. doi: 10.1016/j.jphs.2018.02.001

50. Zhang J, Wang H, Sherbini O, Ling-Lin Pai E, Kang SU, Kwon JS, et al. High-content genome-wide RNAi screen reveals CCR3 as a key mediator of neuronal cell death. *eNeuro.* (2016) 3:1–3. doi: 10.1523/ENEURO.0185-16.2016

51. Kitayama J, Mackay CR, Ponath PD, Springer TA. The C-C chemokine receptor CCR3 participates in stimulation of eosinophil arrest on inflammatory endothelium in shear flow. *J Clin Invest.* (1998) 101:2017–24. doi: 10.1172/JCI2688

52. Gauvreau GM, Fitzgerald JM, Boulet LP, Watson RM, Hui L, Villineuve H. The effects of a CCR3 inhibitor, AXP1275, on allergen-induced airway responses in adults with mild-to-moderate atopic asthma. *Clin Exp Allergy.* (2018) 48:445–51. doi: 10.1111/cea.13114

53. Yoshie O, Matsushima K. CCR4 and its ligands: from bench to bedside. *Int Immunol.* (2015) 27:11–20. doi: 10.1093/intimm/dxu079

54. Kumai T, Nagato T, Kobayashi H, Komabayashi Y, Ueda S, Kishibe K, et al. CCL17 and CCL22/CCR4 signaling is a strong candidate for novel targeted therapy against nasal natural killer/T-cell lymphoma. *Cancer Immunol Immunother.* (2015) 64:697–705. doi: 10.1007/s00262-015-1675-7

55. Butterfield JH, Leiferman KM, Abrams J, Silver JE, Bower J, Gonchoroff N, et al. Elevated serum levels of interleukin-5 in patients with the syndrome of episodic angioedema and eosinophilia. *Blood.* (1992) 79:688–92. doi: 10.1182/blood.V79.3.688.688

56. Tan S, Shan Y, Lin Y, Liao S, Zhang B, Zeng Q, et al. Neutralization of interleukin-9 ameliorates experimental stroke by repairing the blood-brain barrier via down-regulation of astrocyte-derived vascular endothelial growth factor-A. *FASEB J.* (2019) 33:4376–87. doi: 10.1096/fj.20180 1595RR

57. Kouro T, Takatsu K. IL-5- and eosinophil-mediated inflammation: from discovery to therapy. *Int Immunol.* (2009) 21:1303–9. doi: 10.1093/intimm/dxp102

58. Roufosse F. Targeting the interleukin-5 pathway for treatment of eosinophilic conditions other than asthma. *Front Med.* (2018) 5:49. doi: 10.3389/fmed.2018.00049

59. Jia L, Wang Y, Li J, Li S, Zhang Y, Shen J, et al. Detection of IL-9 producing T cells in the PBMCs of allergic asthmatic patients. *BMC Immunol.* (2017) 18:38. doi: 10.1186/s12865-017-0220-1

60. Pflanz S, Timans JC, Cheung J, Rosales R, Kanzler H, Gilbert J. IL-27, a heterodimeric cytokine composed of EBI3 and p28 protein, induces proliferation of naive CD4+ T cells. *Immunity.* (2002) 16:779–90. doi: 10.1016/S1074-7613(02)00324-2

61. Takeda A, Hamano S, Yamanaka A, Hanada T, Ishibashi T, Mak TW, et al. Cutting edge: role of IL-27/WSX-1 signaling for induction of T-bet through activation of STAT1 during initial Th1 commitment. *J Immunol.* (2003) 170:4886–90. doi: 10.4049/jimmunol.170.10.4886

62. Lucas S, Ghilardi N, Li J, de Sauvage FJ. IL-27 regulates IL-12 responsiveness of naive CD4+ T cells through Stat1-dependent and -independent mechanisms. *Proc Natl Acad Sci USA.* (2003) 100:15047–52. doi: 10.1073/pnas.2536517100

63. Batten M, Ghilardi N. The biology and therapeutic potential of interleukin 27. *J Mol Med.* (2007) 85:661–72. doi: 10.1007/s00109-007-0164-7

64. Tait Wojno ED, Hunter CA, Stumhofer JS. The immunobiology of the interleukin-12 family: room for discovery. *Immunity.* (2019) 50:851–70. doi: 10.1016/j.immuni.2019.03.011

65. Batten M, Li J, Yi S, Kljavin NM, Danilenko DM, Lucas S, et al. Interleukin 27 limits autoimmune encephalomyelitis by suppressing the development of interleukin 17-producing T cells. *Nat Immunol.* (2006) 7:929–36. doi: 10.1038/ni1375

66. Stumhofer JS, Laurence A, Wilson EH, Huang E, Tato CM, Johnson LM. Interleukin 27 negatively regulates the development of interleukin 17-producing T helper cells during chronic inflammation of the central nervous system. *Nat Immunol.* (2006) 7:937–45. doi: 10.1038/ni1376

67. Diveu C, McGeachy MJ, Cua DJ. Cytokines that regulate autoimmunity. *Curr Opin Immunol.* (2008) 20:663–8. doi: 10.1016/j.coi.2008.09.003

68. Trinchieri G, Pflanz S, Kastelein RA. The IL-12 family of heterodimeric cytokines: new players in the regulation of T cell responses. *Immunity.* (2003) 19:641–4. doi: 10.1016/S1074-7613(03)00296-6

69. Zhao X, Ting SM, Liu CH, Sun G, Kruzel M, Roy-O'Reilly M, et al. Neutrophil polarization by IL-27 as a therapeutic target for intracerebral hemorrhage. *Nat Commun.* (2017) 8:602. doi: 10.1038/s41467-017-00770-7