Dissertation der

Graduate School of Systemic Neurosciences

der Ludwig-Maximilians-Universität München

aus dem

Institut für Schaganfall- und Demenzforschung

der Ludwig-Maximilians-Universität München

The role of blood-based biomarkers in ischemic stroke

Steffen Tiedt

23.10.2017

Graduate School of
Systemic Neurosciences
LMU Munich

Supervisor:	Prof. Dr. Martin Dichgans
Second reviewer:	Prof. Dr. Dieter Edbauer
Date of defense:	06.03.2018

Melina und

meinen Eltern

Table of Contents

1. List of abbreviations

This list of abbreviations is independent of abbreviation lists in the research articles.

AF	Atrial fibrillation
Ago	Argonaute
BNP	Brain natriuretic peptide
CCS	Causative classification of stroke
CT	Computed tomography
ESUS	Embolic stroke of undetermined source
GWAS	Genome-wide association study
HC	Healthy control subject
HS	Hemorrhagic stroke
IS	Ischemic stroke
miRNA	microRNA
mRNA	messenger RNA
MRI	Magnetic resonance imaging
MT	Mechanical thrombectomy
Nt	Nucleotide
rt-PA	Recombinant tissue plasminogen activator
SNP	Single-nucleotide polymorphism
TBI	Traumatic brain injury
TIA	Transient ischemic attack
TOAST	The Trial of Organization 10172 in Acute Stroke Treatment

2. Abstract

Stroke is a major cause of death and disability around the globe. Acute treatment options in defined time windows and secondary prevention are the mainstays for improving outcome and reducing recurrence. However, the identification of patients eligible for treatments by current diagnostic tests remains insufficient. Currently, there are no blood-based or genetic biomarkers supporting clinical decision-making in ischemic stroke (IS). Consequently, treatment decisions are based on delayed hospital-based neuroimaging in the acute phase and the integration of clinical and technical findings into classification systems for the guidance of secondary prevention. Circulating microRNAs (miRNAs) exhibit cell-type specific expression patterns and high stability and have therefore been proposed as disease markers. We aimed to 1) identify circulating miRNAs associated with acute IS and explore their utility as an early diagnostic marker, 2) determine the agreement between two established classification systems for stroke etiology, and 3) identify genetic variants associated with IS and its etiological subtypes.

First, using RNA sequencing and qRT-PCR in three independent samples (N=492) we established a set of three circulating miRNAs, miR-125a-5p, miR-125b-5p and miR-143-3p, which we found to be elevated in the acute phase of IS. Their combination showed unprecedented diagnostic utility and outperformed routine multimodal CT. Addressing our second aim we found the agreement between the two classification systems CCS and TOAST to be moderate ($\kappa = 0.59$). In either system, the etiology could not be determined in more than 25 % of the patients. Third, studying the association of single-nuleotide polymorphisms with IS over the whole genome in 37.893 IS cases and 397.209 controls, we identified a novel association of large artery stroke with a locus near *TSPAN2* and confirmed four previously established loci, *PITX2* and *ZFHX3* for cardioembolic stroke, *HDAC9* for large artery stroke, and 12q24 near *ALDH2* for small artery stroke.

Taken together, these studies reflect the potential of blood-based biomarkers to support the diagnosis of IS, the insufficiency of current diagnostic tests to determine stroke etiology in all patients, and genetic differences between stroke subtypes. The identification of blood-based biomarkers might improve stroke diagnostics not only for the differentiation of stroke subtypes, but also of hemorrhagic from ischemic stroke.

3. Introduction

3.1. Ischemic stroke

Ischemic stroke (IS) is defined as an episode of neurological dysfunction caused by infarction of the central nervous system due to the narrowing or occlusion of a brain supplying artery as evidenced by either objectives like pathology and imaging or duration of symptoms of more than 24 hours.[1] Stroke, 85 % being ischemic, is the second leading cause of death and a major contributor to long-term disability and dementia.[2] Hypertension, diabetes mellitus, cigarette smoking, and atrial fibrillation (AF) among others have been identified as important risk factors.[3] IS is mostly caused by an embolus occluding a more distal vessel with subsequent activation of the coagulation cascade, platelet aggregation and endothelial dysfunction.[4] Due to the brain's high intrinsic metabolic activity and high concentrations of the excitatory neurotransmitter glutamate, it has a superior demand in oxygen and glucose and is therefore particularly vulnerable to ischemia.[5] The decreased delivery of oxygen and glucose to the brain leads to failure of ion pumps as well as reuptake mechanisms. Subsequent excitotoxicity and cortical spreading depolarizations eventually result in neuronal and glial injury.[5,6] Acute and chronic activation as well as subacute depression of the immune system make stroke not only a disorder restricted to the brain, but affecting the whole organism.

Systemically administered recombinant tissue plasminogen activator (rt-PA) represents the only approved drug for patients with IS in the acute phase. Catalyzing the conversion of plasminogen to plasmin, which in turn degrades fibrin, rt-PA is intended to resolve the blood clot and eventually lead to reperfusion. Recently, it was shown that patients with occlusions of arteries of the proximal anterior circulation further benefit from interventional endovascular mechanical thrombectomy (MT) compared to standard medical care.[7]

3.1.1. Genetics of IS

In the general population, prior stroke in a first-degree relative increases stroke risk by 30 %.[8] Hereditary factors contribute to stroke risk through several distinct mechanisms: 1) rare single gene mutations causing familial disorders with stroke being the primary manifestation (e.g. cerebral autosomal dominant arteriopathy with subcortical infarcts and

leukoencephalopathy (CADASIL)), 2) single gene mutations causing familial disorders with stroke being one of many manifestations (e.g. Marfan syndrome), 3) common variants of genetic polymorphisms associated with an increase in stroke risk (e.g. rs11984041 at the *HDAC9* locus for large artery atherosclerosis-related stroke), and 4) genetic variability associated with risk factors for IS, e.g. AF, diabetes mellitus, and hypertension (reviewed by Boehme et al.)[9].

Studying associations of genetic variants in the whole genome with a distinct phenotype, known as genome-wide association study (GWAS), yields the potential to identify loci involved in the pathogenesis of a disease and use these findings to accelerate development of diagnostics and treatment. Up to date, the loci *PITX2,*[10] *ZFHX3,*[11] *HDAC9,*[12] and *12q24.12*[13] have been associated with IS using GWAS. Studying more than 300.000 genetic variants, so-called single-nucleotide polymorphisms (SNP), in three independent samples of Icelandic and European origin with more than 6.000 patients with IS and 30.000 control subjects, Gretarsdottir and colleagues associated two SNPs close to the *PITX2* locus, known to be a regulator of the development of the sinoatrial node, with cardioembolic stroke.[10] Next, three of the top ten SNPs from this study were investigated in eight independent samples from Iceland, Norway, Sweden, Germany, the UK and the United States. The sequence variant rs7193343, intronically located at the *ZFHX3* locus on chromosome 16q22, was found to be significantly associated with AF and cardioembolic IS.[11] The first locus associated with large vessel stroke, *HDAC9*, was identified in three independent samples of European ancestry by collaboration within the International Stroke Genetics Consortium.[12] The METASTROKE consortium identified *12q24.12*, the first locus being associated with all IS rather than a specific stroke subtype.[13] Studying a larger number of patients with less phenotypic heterogeneity may lead to the identification of even more loci associated with IS subtypes.

3.1.2. Stroke etiology

The decision for secondary prevention of IS patients depends on the determination of the etiology of the individual patient's IS. Accordingly, classification systems have been developed. The Trial of Organization 10172 in Acute Stroke Treatment (TOAST) criteria distinguish between large-artery atherosclerosis, defined as stenosis of more than 50 % or occlusion of a major brain or branch cortical artery, cardioembolism, an embolus from the

heart resulting e.g. from AF, small-artery occlusion, characterized by smaller subcortical or brainstem infarcts, stroke of other determined etiology, e.g. arterial dissection, and undetermined etiology.[14] Using this classification system, approximately 25 % of IS are categorized as large artery atherosclerosis, 30 % as cardioembolic, 10 % as small-artery occlusion and 5 % as other defined etiology. In approximately 30 % of IS patients the etiology cannot be determined.[15] The more recent causative classification of stroke (CCS), a web-based algorithm, has adapted the TOAST criteria and assigned categorical probabilities. In a limited number of patients, this tool minimized the classification into the category "undetermined" to 4 %.[16] Both, the agreement between TOAST and CCS and the classification performance of CCS, have not been determined in a larger sample.

Determination of stroke etiology subsequently guides secondary prevention. In a simplified way, patients with IS due to AF, responsible for the majority of cardioembolic events, are treated with anticoagulating agents, whereas patients with IS due to non-cardioembolic causes are mostly treated with antiplatelet agents.

3.1.3. Diagnostic measures of IS

To put this work in a more general framework of diagnostics in IS and to be able to later identify novel prospects for blood-based biomarker in this context I will here introduce some stroke-specific diagnostic questions whose answers would guide therapeutic decision making.

In order to identify patients eligible for rt-PA: are currently applied tests for the identification of patients with hemorrhagic stroke (HS) fast and reliable? Computed tomography (CT) is the method of choice to detect intracranial hemorrhages. Its sensitivity for this question approximates 100 %, its acquisition within the hospital is cheap and fast. Magnetic resonance imaging (MRI) offers equal sensitivity and depending on the chosen modalities the possibility to gain more information about the potentially underlying disease, but should only be applied in the acute phase if, compared to CT, causing no delay.[17,18] Both techniques require the patient's admission to the hospital. Currently, it is not possible to reliably rule out intracranial bleeding in the prehospital setting.

In order to identify patients eligible for MT: are currently applied tests for the diagnosis of the occlusion of an artery of the proximal anterior or posterior circulation fast and reliable? CT-angiography is the method of choice to visualize supraaortic vessels in patients with suspected acute IS and therefore to identify occlusions of arteries of the proximal cerebral circulation with a sensitivity higher than 90 %.[17,18] Given the restriction of MT to specialized stroke centers, the prehospital decision process includes the following question: should a patient with suspected acute IS be transferred to the nearest local stroke unit with available CT and subsequent rt-PA application and in case of a proximal arterial occlusion be transferred a second time to a specialized stroke center for MT or should the patient be directly transferred to the specialized stroke center with a resulting delay in systemic thrombolysis and uncertainty if MT is needed at all? Currently, it is not possible to identify patients with an occlusion of an artery of the proximal anterior circulation in the prehospital setting.

In order to guide secondary prevention: are currently applied tests sensitive and specific enough to distinguish patients with IS from patients with other diseases causing acute neurological dysfunction? MRI offers > 90 % sensitivity and high specificity for the identification of cerebral infarction and is also more accurate in describing size and localization of infarcted areas compared to CT.[17] Clinically suspected but MRI negative ischemic strokes are mostly assumed to be brainstem or lacunar infarctions smaller than the typical 3 or 5 mm MRI sections (Stroke 2008, Sylaja). However, MRI is an expensive technique, not available for every patient, especially not in third-world countries.

In order to guide secondary prevention: are currently applied tests sensitive and specific enough to identify stroke etiology? Current guidelines for secondary prevention mainly pose two questions to guide treatment decisions. 1) Does the patient have a stenosis of more than 70 % based on the NASCET criteria of the internal carotid artery? This question can accurately be answered by duplex-/doppler-sonography and CT-angiography. 2) Did the cerebral vessel occlusion result from cardioembolism or more specifically AF? Electrocardiography shows almost 100 % sensitivity for AF. However, many AF cases are paroxysmal and subclinical, thereby complicating their detection. Consequently, longer monitoring after IS has been associated with higher AF detection rate.[19] Although all trials guiding our current treatments have only included patients with ECG proven atrial fibrillation,

recent studies suggest that AF could also be an epiphenomenon and not necessarily the cause. For example, no temporal association was found between IS onset and subclinical AF episodes in patients with implantable pacemakers and defibrillators.[20] Therefore, it has been increasingly questioned how and when the embolus actually evolves in patients with AF. Understanding the underlying pathophysiology may lead to new (blood-based) diagnostic possibilities.

3.1.4. Blood-based biomarkers of IS

I will here focus on the value of blood-based biomarkers to address some of the questions raised in the previous paragraph. Currently, none of these biomarkers is used in clinical routine.

What is the value of blood-based biomarkers to differentiate patients with HS from patients with IS? Presumably due to the immediate mechanical astrocytic injury, the serum levels of the intermediate filament glial fibrillary acidic protein (GFAP) increase very early after intracerebral hemorrhage.[21] Showing also increased levels in the subacute phase of IS, GFAP was demonstrated to have highest value in differentiating HS from IS up to six hours after symptom onset. Although the proposed sensitivity (39 – 91 %) and specificity (96 - 100 %) of GFAP show potential,[22-25] small sample sizes and variable cut-offs between studies have prevented its translation into clinical and preclinical routine.

What is the value of blood-based biomarkers for the diagnosis of the occlusion of an artery of the proximal anterior or posterior circulation? Neither single biomarkers nor biomarker sets have been studied regarding their value in detecting these occlusions yet. Available point-of-care-testing of blood-based biomarkers in the preclinical setting could result in more accurate and faster decisions, e.g. about whether the patient should be transferred to a stroke center with endovascular thrombectomy service or to a local stroke unit for immediate thrombolysis.

What is the value of blood-based biomarkers to distinguish patients with IS from patients with other diseases causing acute neurological dysfunction? To address this question, previous studies mainly focused on targeted protein analysis. Both, single cerebral proteins, e.g. neuron-specific enolase,[26] GFAP,[27] or S100 calcium binding protein B,[28] and

combinations of proteins and clinical parameters,[29] show limited diagnostic accuracy in the acute phase.

What is the value of blood-based biomarkers to identify stroke etiology? Differentiating cardioembolic from non-cardioembolic stroke remains the key question in guiding secondary prevention. Observational data identified the brain natriuretic peptide (BNP) and the inactive prohormone N-terminal fragment as surrogate marker for cardioembolic stroke, in particular following heart failure and AF,[30-32] with a sensitivity of 90 % and a specificitiy of 80 %. Although the probability of cardioembolism correlates with BNP serum levels, varying thresholds and assays between studies have prevented its widespread application in clinical routine.

3.2. MicroRNAs

MicroRNAs (miRNAs) have a length of approximately 19-24 nucleotides (nt), do not code for proteins, but exert a posttranscriptional regulatory function. Accordingly, they are categorized as short, non-coding, regulatory RNAs. Canonical multilevel biogenesis is initiated by RNA polymerase II-transcription into primary miRNAs and subsequent liberation of a stem-loop precursor miRNA of 60-70 nt length through endoribonucleolytic cleavage. Export to the cytoplasm is followed by a second endoribonucleolytic reaction into a ~ 22 nt RNA duplex consisting of the guide strand (the mature miRNA) and a quickly degraded passenger strand (reviewed by Bartel, Mendell and Olson, and Yates et al.)[33-35].

Mature miRNAs exert their posttranscriptional regulatory function as a specificity determinant to messenger RNAs (mRNAs) for Argonaute (Ago) proteins within the miRNA-induced silencing complex. While fully complementary mRNAs are cleaved, partially complementary targets are subject to a combination of translational repression, deadenylation, decapping and 5'-to-3'-degradation. The recognition of partially complementary mRNAs primarily depends on the seed sequence (nt positions 2 to 8) of a miRNA. Current evidence suggests that miRNAs exert their function by subtly regulating the expression of many targets leading to a stronger network effect and eventually phenotypic changes (reviewed by Mendell and Olson, Yates et al., and Jonas and Izzauralde)[34-36]. miRNAs have been shown to be involved in numerous pathophysiological processes and therefore

suggested to be potential drug targets, e.g. for seizures,[37] obesity-related glucose metabolism,[38] myocardial disease,[39] and atherosclerosis.[40]

The detection of miRNAs in the circulation[41] has created the possibility to study the value of miRNAs as biomarkers for disease.[42] Being bound to RNA-binding proteins, e.g. Ago2,[43] and lipoproteins,[44] or present within extracellular vesicles, e.g. exosomes,[45] miRNAs exert remarkable stability within this RNase-rich compartment, but also during storage and freeze-thaw cycles.[41] Their use as diagnostic and prognostic biomarkers has been studied in multiple disease entities including cardiovascular and malignant diseases (reviewed by Creemers et al., and He et al.)[46,47]. Extracellular vesicles are also able to fuse with recipient cells enabling intercellular signaling of miRNAs.[48] Thus, vesicle-mediated transfer of miRNAs between cells could impact on pathophysiological processes during disease.[49,50]

Showing cell-type specific expression patterns, circulating miRNAs may reflect pathophysiological processes of IS including platelet aggregation, neuronal and glial injury.[51,52]

4. Research articles

4.1. RNA-Seq Identifies Circulating miR-125a-5p, miR-125b-5p and miR-143-3p as Potential Biomarkers for Acute Ischemic Stroke

Up to 50 % of patients with acute ischemic stroke (IS) lack abnormalities on admission computed tomography (CT) scans. The diagnostic value of single protein blood-based biomarkers in the acute phase of IS is limited. The cell-type specific expression pattern of microRNAs (miRNAs) and their high stability in blood promise potential diagnostic utility for multiple disease entities. We hypothesized that a set of miRNAs might better reflect the different pathophysiological processes of IS such as platelet aggregation, endothelial dysfunction, neuronal and glial injury, and immunological reactions compared to a single biomarker.

Using RNA sequencing and qRT-PCR in three independent patient samples (N=260 IS patients) we established a set of three circulating miRNAs, miR-125a-5p, miR-125b-5p and miR-143-3p, which were elevated in the acute phase of IS compared to healthy control subjects (HC) or patients with transient ischemic attacks (TIA). Longitudinal analysis of expression levels up to 90 days after IS revealed continuously elevated levels of miR-125a-5p, while levels of miR-125b-5p and miR-143-3p normalized to control levels starting at day two. The levels of these miRNAs were unaffected by chemical hypoxia in N2a cells and in experimental stroke models, but depended on platelet numbers in spike-in experiments suggesting a relation to thrombus formation or platelet aggregation. Showing unprecedented diagnostic utility for a blood-based biomarker in the acute phase of IS, this combination of miRNAs outperformed routine multimodal CT.

These findings encourage larger multi-center studies to determine the full potential of these miRNAs in guiding clinical decision making.

Author contribution: design of experiments, recruitment of patients and healthy control subjects, performing the experiments, analyses of the experiments, interpretion of the results, and writing the manuscript (please see section 10 for further details).

RNA-Seq Identifies Circulating miR-125a-5p, miR-125b-5p, and miR-143-3p as Potential Biomarkers for Acute Ischemic Stroke

Steffen Tiedt, Matthias Prestel, Rainer Malik, Nicola Schieferdecker, Marco Duering, Veronika Kautzky, Ivelina Stoycheva, Julia Böck, Bernd H. Northoff, Matthias Klein, Franziska Dorn, Knut Krohn, Daniel Teupser, Arthur Liesz, Nikolaus Plesnila, Lesca Miriam Holdt, Martin Dichgans

Rationale: **Currently, there are no blood-based biomarkers with clinical utility for acute ischemic stroke (IS). MicroRNAs show promise as disease markers because of their cell type–specific expression patterns and stability in peripheral blood.**

Objective: **To identify circulating microRNAs associated with acute IS, determine their temporal course up to 90 days post-stroke, and explore their utility as an early diagnostic marker.**

Methods and Results: **We used RNA sequencing to study expression changes of circulating microRNAs in a discovery sample of 20 patients with IS and 20 matched healthy control subjects. We further applied quantitative real-time polymerase chain reaction in independent samples for validation (40 patients with IS and 40 matched controls), replication (200 patients with IS, 100 healthy control subjects), and in 72 patients with transient ischemic attacks. Sampling of patient plasma was done immediately upon hospital arrival. We identified, validated, and replicated 3 differentially expressed microRNAs, which were upregulated in patients with IS compared with both healthy control subjects (miR-125a-5p [1.8-fold; $P=1.5\times10^{-6}$], miR-125b-5p [2.5-fold; $P=5.6\times10^{-6}$], and miR-143-3p [4.8-fold; $P=7.8\times10^{-9}$]) and patients with transient ischemic attack (miR-125a-5p: $P=0.003$; miR-125b-5p: $P=0.003$; miR-143-3p: $P=0.005$). Longitudinal analysis of expression levels up to 90 days after stroke revealed a normalization to control levels for miR-125b-5p and miR-143-3p starting at day 2 while miR-125a-5p remained elevated. Levels of all 3 microRNAs depended on platelet numbers in a platelet spike-in experiment but were unaffected by chemical hypoxia in Neuro2a cells and in experimental stroke models. In a random forest classification, miR-125a-5p, miR-125b-5p, and miR-143-3p differentiated between healthy control subjects and patients with IS with an area under the curve of 0.90 (sensitivity: 85.6%; specificity: 76.3%), which was superior to multimodal cranial computed tomography obtained for routine diagnostics (sensitivity: 72.5%) and previously reported biomarkers of acute IS (neuron-specific enolase: area under the curve=0.69; interleukin 6: area under the curve=0.82).**

Conclusions: **A set of circulating microRNAs (miR-125a-5p, miR-125b-5p, and miR-143-3p) associates with acute IS and might have clinical utility as an early diagnostic marker. (*Circ Res*. 2017;121:970-980. DOI: 10.1161/CIRCRESAHA.117.311572.)**

Key Words: biomarkers ■ microRNAs ■ stroke

Stroke is the leading cause of long-term disability and the second most common cause of death worldwide with ≈80% of cases attributed to ischemia.[1] Currently, the diagnosis of ischemic stroke (IS) on hospital arrival is often reduced to the exclusion of hemorrhagic stroke by computed tomography (CT), and 40% to 50% of all acute IS cases lack abnormalities on admission CT scans.[2,3] The diagnostic value of blood biomarkers in acute IS is still limited.[4] Previous attempts

Original received June 28, 2017; revision received July 14, 2017; accepted July 18, 2017. In June 2017, the average time from submission to first decision for all original research papers submitted to *Circulation Research* was 12.45 days.

From the Institute for Stroke and Dementia Research, Klinikum der Universität München (S.T., M.P., R.M., N.S., M.D., V.K., I.S., J.B., A.L., N.P., M.D.), Graduate School of Systemic Neurosciences (S.T.), Institute of Laboratory Medicine, Klinikum der Universität München (B.H.N., D.T., L.M.H.), Department of Neurology, Klinikum der Universität München (M.K.), and Department of Neuroradiology, Klinikum der Universität München (F.D.), Ludwig-Maximilians-Universität LMU, Germany; Munich Cluster for Systems Neurology (SyNergy), Germany (S.T., A.L., N.P., M.D.); and Interdisciplinary Center for Clinical Research (IZKF), University of Leipzig, Germany (K.K.).

The online-only Data Supplement is available with this article at http://circres.ahajournals.org/lookup/suppl/doi:10.1161/CIRCRESAHA.117.311572/-/DC1.

Correspondence to Martin Dichgans, MD, Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität LMU, Feodor-Lynen-Straße 17, 81377 Munich, Germany. E-mail martin.dichgans@med.uni-muenchen.de

Novelty and Significance

What Is Known?

- Up to 50% of patients with acute ischemic stroke (IS) lack abnormalities on admission computed tomographic scans, currently the gold standard for diagnostic evaluation.
- Protein markers of brain injury show limited diagnostic value as blood-based biomarkers in the acute phase of IS.
- MicroRNAs exhibit cell type–specific expression patterns and high stability in peripheral blood.

What New Information Does This Article Contribute?

- Expression levels of circulating miR-125a-5p, miR-125b-5p, and miR-143-3p are elevated in the acute phase of IS.
- Human platelets are a major source of these microRNAs, whereas neuronal injury does not contribute to their levels in experimental paradigms.
- The combination of all 3 microRNAs shows unprecedented diagnostic utility for a blood-based biomarker in the acute phase of IS and outperforms routine multimodal computed tomography.

Previous studies have demonstrated the limited diagnostic value of single protein markers in the acute phase of IS. Because of their cell type–specific expression pattern, microRNAs show promise as diagnostic markers for multiple disease entities. We hypothesized that a set of microRNAs might better reflect pathophysiological processes of IS, such as platelet aggregation, endothelial dysfunction, neuronal and glial injury, and immunologic reactions. Using RNA sequencing and quantitative real-time polymerase chain reaction in 3 independent patient samples, we established a set of 3 circulating microRNAs, which was elevated in the acute phase of IS. Their levels were unaffected in 2 experimental stroke models and after neuronal injury in vitro but depended on platelet numbers in spike-in experiments, suggesting a relationship to platelet aggregation or thrombus formation. The combination of all 3 microRNAs showed unprecedented diagnostic utility for a blood-based biomarker in the acute phase of IS and outperformed routine multimodal computed tomography. Our findings encourage studies in even larger patient populations to determine the full potential of these microRNAs in guiding clinical decision making.

Nonstandard Abbreviations and Acronyms

AUC	area under the curve
CT	computed tomography
HC	healthy control subject
IS	ischemic stroke
TIA	transient ischemic attack

to define markers of IS mostly focused on proteins, such as neuron-specific enolase,[5] glial fibrillary acidic protein,[6] S100 calcium-binding protein B,[7] or interleukin 6.[8] However, these markers all either show limited specificity or substantial delays in response to acute brain injury.[4]

Editorial, see p 905
In This Issue, see p 893
Meet the First Author, see p 894

MicroRNAs represent a class of small noncoding RNAs, which have widely been studied for their intracellular role in post-transcriptional modification.[9] They can also be released into the circulation, where they are present in extracellular vesicles or bound to protein complexes.[10,11] Because of their stability in peripheral blood and cell type–specific expression patterns, microRNAs show promise as diagnostic and prognostic markers for multiple disease entities.[12–14]

IS might impact on expression levels of circulating microRNAs through various simultaneously occurring pathophysiological processes, including platelet aggregation, endothelial dysfunction, neuronal and glial injury, as well as immunologic reactions.[15–19] Hence, a signature of microRNAs resulting from different pathophysiological events could have higher diagnostic accuracy than single biomarkers. However, there is no analysis of the complete profile of microRNAs in the acute phase of IS immediately after hospital arrival.

The aim of the current study was to (1) identify differentially regulated circulating microRNAs after acute IS using RNA sequencing, (2) validate and replicate findings in 2 independent samples, (3) characterize their temporal profile during hospitalization and up to 90 days after IS, and (4) investigate their potential utility as a biomarker for acute IS. We further obtained data on the cellular source and the extracellular mode of transportation of differentially expressed microRNAs.

Methods

An expanded Methods section is available in the Online Data Supplement.

Study Population

Patients with suspected IS were recruited within 24 hours of symptom onset through the emergency department of a tertiary level university hospital. Blood samples (nonfasting) were collected before the receipt of any medication. All patients had a final diagnosis of IS as defined by an acute focal neurological deficit in combination with a diffusion-weighted imaging–positive lesion on magnetic resonance imaging or a new lesion on a delayed CT scan. Transient ischemic attack (TIA) was defined as acute onset transient symptoms (lasting <24 hours) without evidence of infarction on neuroimaging.[20] The discovery sample included 20 patients with IS and 20 healthy control subjects (HCs). Validation was done in 40 patients with IS and 40 HCs. The replication sample consisted of 200 patients with IS and 100 HCs. HCs from the discovery and validation sample were matched to patients with IS for age, sex, hypertension, smoking history, hypercholesterolemia, obesity, diabetes mellitus, family history, and use of antiplatelet therapy. We further studied 72 patients with TIA.

Discovery Analysis Using RNA Sequencing

RNA <1000 nucleotides was isolated from 2 plasma aliquots of 200 µL using the miRCURY RNA Isolation Kit—Biofluids (Exiqon; Online Data Supplement). Reverse transcription was performed with the miScript Reverse Transcription kit (Qiagen) according to the manufacturer's protocols in a 10 µL reaction. A total of 500 ng of RNA prepared from the plasma samples was used in the small RNA protocol with the TruSeq Small RNA sample prepkit v2 (Illumina) following the instructions of the manufacturer. For details, see Braun

et al.[21] The barcoded libraries were size restricted between 140 and 165 base pairs for additional enrichment of microRNAs, purified, and quantified using the Library Quantification Kit-Illumina/Universal (KAPA Biosystems, Woburn, MA) according to the instructions of the manufacturer. Libraries were pooled at a concentration of 10 nmol/L for cluster generation using an Illumina cBot. Sequencing of 50 base pairs was performed with an Illumina HighScan-SQ sequencer at the sequencing core facility of the IZKF Leipzig (Faculty of Medicine, University Leipzig) using version 3 chemistry and flow cell following the instructions of the manufacturer.

RNA Sequencing Data Analysis

Demultiplexing of raw reads, adapter trimming, and quality filtering were done according to Stokowy et al.[22] Further analysis was based on bowtie2 alignment to the sequence of the human genome (hg19) allowing 1 mismatch and alignments to multiple reference targets. MicroRNA sequences were named in concordance with miRBase 20. R packages DESeq2[23] and EdgeR[24] were used to normalize counts and calculate differential expression of microRNAs.

Results

Characteristics of the Study Sample

Patients and control subjects included in the discovery samples were matched with respect to all variables except for glucose levels. The validation samples were matched for demographic and vascular risk factors, as well as previous use of antiplatelet medication. The mean times from symptom onset to blood sampling in the discovery, validation, and replication sample were 3.9, 6.1, and 5.0 hours, respectively (Table 1). Stroke pathogeneses in the 260 patients with IS included large-artery atherosclerosis (23.5%), cardioembolism (30.4%), small-vessel occlusion (6.9%), and arterial dissection (2.3%). In 36.9% of the patients, the pathogenesis remained undetermined.

Profiling of Circulating MicroRNAs in Acute Stroke

To determine the profile of circulating microRNAs associated with IS in the acute phase, we first performed RNA sequencing in the discovery sample (Figure 1A). Sequencing of small RNAs isolated from platelet-poor plasma revealed a median read count of 8.1 million both in patients with IS (interquartile range, 6.8–8.5 million) and in HCs (interquartile range, 5.6–9.5 million; for individual sample characteristics, see Online Table I). Overall, we detected 532 mature microRNAs with a minimum of 5 reads in all 40 samples. To identify differentially expressed microRNAs, we applied 2 independent strategies for the normalization of read counts, DESeq2[23] and EdgeR.[24] Thirty-two mature microRNAs showed a false discovery rate–adjusted $P<0.05$ in both normalized data sets (Figure 1B; Table 2). This selection was independent of microRNA abundance (Online Figure I). Twenty-eight of these microRNAs were downregulated and 4 were upregulated in patients with IS compared with HCs. The microRNA reads per sample (raw and normalized) and microRNA statistics (EdgeR and DESeq2) are presented in Online Table IIa and IIb, respectively.

Validation of Differentially Expressed MicroRNAs

Next, we examined these 32 microRNAs in an independent validation sample using quantitative real-time polymerase chain reactions (Figure 1A). Twenty-two microRNAs passed prespecified quality criteria (Table 2). Six of these microRNAs were differentially regulated after stroke ($P<0.00227$ after Bonferroni correction). Three of them showed consistent directionality with the RNA sequencing results in the discovery sample. They were all upregulated in patients with IS compared with HCs: miR-143-3p (4.77±0.70-fold, mean±SEM; $P=7.8\times10^{-9}$), miR-125b-5p (2.54±0.36-fold; $P=5.6\times10^{-6}$), and miR-125a-5p (1.80±0.16-fold; $P=1.5\times10^{-6}$; Figure 2A; for other microRNAs, see Online Figure II). The results remained stable when using unadjusted raw values and when excluding subjects with previous major cardiovascular events (Online Figure III). The results were also stable when comparing microRNA copy numbers per microliter plasma from HCs and patients with IS from the combined discovery and validation sample (Online Figure IVA and IVB; for additional control experiments, see Online Figures V and VI).

Replication of miR-125a-5p, miR-125b-5p, and miR-143-3p

To further solidify these results, we next examined circulating levels of miR-125a-5p, miR-125b-5b, and miR-143-3p in an independent sample of 200 patients with IS and 100 HCs with additional adjustment for covariates (Figure 1A). Relevant covariates were identified by backward stepwise regression analysis and included age, sex, hypertension, major cardiovascular events, antiplatelet therapy, and usage of calcium channel blockers, as well as creatinine, platelet, red blood cell, and white blood cell counts. All 3 microRNAs showed significantly elevated expression levels in patients with IS compared with HCs (miR-125a-5p: $P=1.1\times10^{-7}$; miR-125b-5p: $P=8.2\times10^{-9}$; miR-143-3p: $P=2.4\times10^{-9}$; Figure 2B). They were also elevated in patients with IS when compared with patients with TIA (miR-125a-5p: $P=0.003$; miR-125b-5p: $P=0.003$; miR-143-3p: $P=0.005$; n=72; Figure 2B).

Temporal Profile of miR-125a-5p, miR-125b-5p, and miR-143-3p During Hospitalization and up to 90 Days After Stroke

To further explore the utility of circulating miR-125a-5p, miR-125b-5p, and miR-143-3p as biomarkers acutely or chronically after stroke, we measured their plasma concentration at the first, second, third, and seventh day of hospitalization and at 90 days post-stroke. Thirty-two patients provided data for at least 4 time points and thus served as our study sample. We found significant changes in expression levels over time for all 3 microRNAs compared with 32 controls (all $P<0.0001$). In all cases, expression levels declined from admission (day 1) to day 2 after stroke (miR-125a-5p: $P=0.0215$; miR-125b-5p: $P<0.0001$; miR-143-3p: $P<0.0001$; Figure 3). Expression levels of miR-125a-5p subsequently increased until day 90 after stroke and were at all time points significantly higher than in HCs. In contrast, expression levels of miR-143-3p remained low after day 2 with no significant difference compared with controls. Likewise, expression levels of miR-125b-5p also showed no significant difference compared with controls starting from day 2. Polynomial regression of microRNA expression levels with exact time points since symptom onset in hours revealed peaks of expression at 2.3 hours (miR-143-3p), 4.6 hours (miR-125b-5p), and 6.5 hours (miR-125a-5p; Online Figure VII). Matched data analysis for changes in expression levels over time revealed significant differences for miR-125b-5p ($P=1.5\times10^{-7}$) and miR-143-3p ($P=3.4\times10^{-6}$). In generalized linear mixed models, miR-125b-5p and miR-143-3p

Table 1. Demographic and Clinical Characteristics of the Discovery, Validation, and Replication Sample

	Discovery Sample		Validation Sample		Replication Sample		
Characteristics	IS Patients	HCs	IS Patients	HCs	IS Patients	HCs	TIA Patients
Demographic characteristics							
Total, n	20	20	40	40	200	100	72
Age, mean (SD), y	74.7 (9.7)	72.7 (10.1)	74.7 (13.8)	69.7 (8.8)	74.1 (13.4)	65.6 (13.4)	74.8 (12.4)
Female, n (%)	12 (60)	10 (50)	18 (45)	24 (60)	87 (43.5)	65 (65)	39 (54.2)
Vascular risk factors, n (%)							
Hypertension	16 (80)	10 (50)	34 (85)	26 (65)	157 (78.9)	35 (35)	54 (75)
Smoking history	8 (40)	9 (45)	20 (50)	12 (30)	74 (40.4)	35 (35)	30 (44.8)
Hypercholesterolemia	5 (25)	3 (15)	11 (27.5)	10 (25)	55 (27.8)	21 (21)	34 (47.2)
Obesity	5 (25)	1 (5)	6 (15)	6 (15)	28 (20.9)	19 (19.2)	13 (21.7)
Diabetes mellitus	4 (20)	0 (0)	4 (10)	2 (5)	36 (18.2)	6 (6)	18 (25.4)
Previous TIA/stroke/MI	0 (0)	0 (0)	13 (32.5)	2 (5)	65 (32.8)	7 (7)	35 (48.6)
Family history	5 (25)	2 (10)	6 (15)	2 (5)	23 (12.8)	9 (9)	9 (13.4)
Laboratory parameters, mean (SD)*							
Glucose, mg/dL	127.8 (30.1)	99.2 (20.1)	129.3 (38.8)	98.5 (20.7)	130.9 (41.6)	102.7 (27.5)	119.8 (34.2)
Creatinine, mg/dL	1.0 (0.2)	1.0 (0.2)	1.1 (0.3)	1.0 (0.2)	1.2 (0.4)	1.0 (0.2)	1.1 (0.4)
hs-CRP, mg/dL	0.7 (1.6)	0.4 (0.4)	0.5 (0.5)	0.2 (0.3)	1.3 (2.6)	0.3 (0.4)	0.7 (2.1)
AST, U/L	24.7 (6.7)	26.2 (5.4)	31.0 (15.9)	27.6 (7.7)	26.4 (11.5)	25.6 (7.1)	23.2 (5.8)
ALT, U/L	21.1 (9.1)	21.3 (6.6)	19.4 (12.6)	23.8 (9.6)	22.5 (16.1)	23.3 (11.7)	20.8 (12.2)
γ-GT, U/L	42.6 (41.6)	26.1 (12.6)	24.1 (7.0)	35.3 (30.9)	45.5 (46.9)	30.8 (25.3)	44.1 (73.8)
Total cholesterol, mg/dL	208.2 (29.7)	222.3 (36.7)	185.8 (46.1)	226.3 (48.3)	185.4 (48.0)	230.7 (49.8)	193.4 (48.6)
Triglycerides, mg/dL	113.5 (40.3)	134.7 (66.0)	104.8 (60.7)	159.1 (71.3)	127.8 (86.1)	173.6 (195.4)	145.0 (73.3)
LDL, mg/dL	134.8 (30.3)	136.9 (23.8)	120.9 (36.3)	144.2 (41.2)	117.9 (39.2)	143.3 (43.3)	123.1 (42.7)
HDL, mg/dL	57.2 (14.1)	67.0 (23.8)	49.6 (12.3)	59.4 (16.8)	49.4 (13.7)	62.3 (16.4)	48.4 (14.6)
WBC count, G/L	7.8 (2.6)	6.7 (1.8)	7.7 (2.2)	6.1 (1.6)	9.3 (3.3)	6.6 (1.6)	7.2 (1.7)
RBC count, T/L	4.7 (0.5)	4.7 (0.4)	4.5 (0.5)	4.7 (0.5)	4.5 (0.6)	4.7 (0.4)	4.5 (0.5)
Hemoglobin, g/dL	14.8 (1.5)	14.2 (0.9)	13.6 (1.6)	14.1 (1.0)	13.7 (1.9)	14.3 (1.1)	13.7 (1.5)
Hematocrit	0.43 (0.04)	0.42 (0.03)	0.40 (0.05)	0.42 (0.03)	0.40 (0.05)	0.42 (0.03)	0.40 (0.04)
Platelets, G/L	210.6 (51.2)	233.9 (60.3)	231.5 (64.8)	240.6 (57.9)	234.8 (79.8)	247.7 (53.8)	217.7 (51.2)
Medication, n (%)							
Statins	5 (25)	3 (15)	11 (27.5)	10 (25)	67 (34.2)	24 (24)	34 (47.9)
ACE I/ARBs	10 (50)	8 (40)	23 (57.5)	15 (37.5)	110 (56.4)	23 (23)	43 (60.6)
β-Blockers	3 (15)	3 (15)	18 (45)	15 (37.5)	91 (46.7)	19 (19)	23 (32.4)
Diuretics	4 (20)	3 (15)	18 (45)	5 (12.5)	72 (36.9)	9 (9)	19 (26.8)
Calcium channel blockers	8 (40)	3 (15)	10 (25)	2 (5)	38 (19.5)	3 (3)	14 (19.7)
Antiplatelet therapy	0 (0)	0 (0)	18 (45)	16 (40)†	66 (33.5)	18 (18)	33 (46.5)
ΔT, mean (SD), h	3.9 (3.6)	n/a	6.1 (4.4)	n/a	5.0 (4.6)	n/a	6.0 (6.4)
Infarct volumes, mean (SD), mL	18.7 (45.3)	n/a	18.6 (38.9)	n/a	30.8 (68.3)	n/a	n/a

γ-GT indicates γ-glutamyl transpeptidase; ΔT, time from symptom onset until hospital arrival; ACE I, angiotensin-converting enzyme inhibitor; ALT, alanine transaminase; ARBs, angiotensin II receptor blockers; AST, aspartate transaminase; HCs, healthy control subjects; HDL, high-density lipoprotein; Hs-CRP, high-sensitivity C-reactive protein; IS, ischemic stroke; LDL, low-density lipoprotein; MI, myocardial infarction; n/a, not available; RBC, red blood cell; TIA, transient ischemic attack; and WBC, white blood cell.

*Blood was sampled in nonfasting condition.

†Sixteen of 40 control subjects in the validation group indicated use of aspirin (acetylsalicylic acid) on a regular basis or at least once within the last 7 days while other antiplatelet drugs were not used. Indications for aspirin use were coronary artery disease (n=4) and pain (n=4) while the indication was not specified in the remaining 8 subjects.

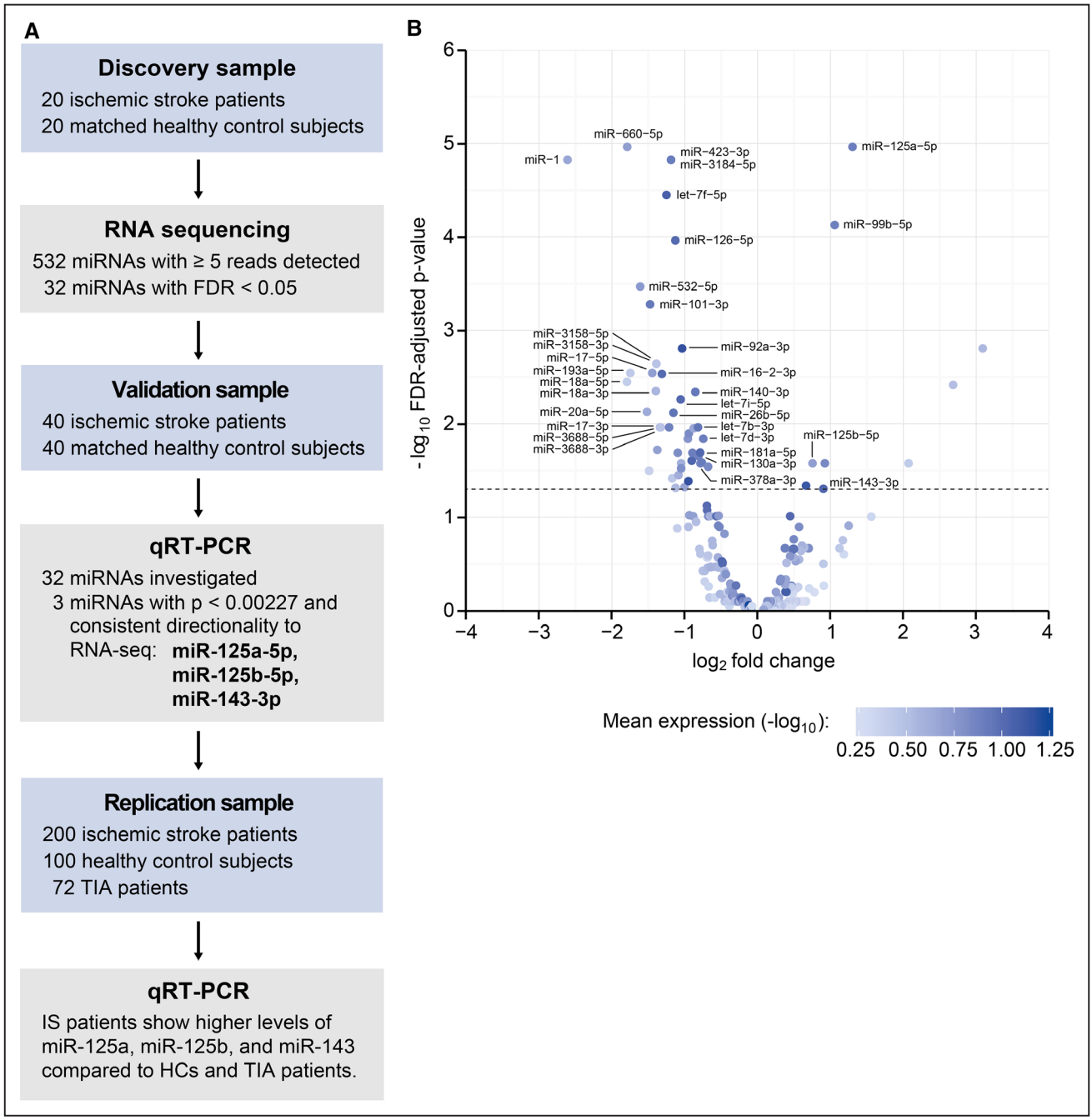


Figure 1. Study profile and RNA sequencing results. A, Workflow illustrating the 3-stage approach involving independent samples for discovery, validation, and replication. **B**, Volcano plot showing RNA sequencing results for the comparison of ischemic stroke patients vs healthy controls. The dashed line indicates a false discovery rate (FDR)–adjusted *P* value of 0.05. Individual microRNAs are displayed by their FDR-adjusted *P* value, and the corresponding fold change with the intensity of blue indicating their mean expression (normalization algorithm EdgeR). MicroRNAs that showed differential expression in both EdgeR and DESeq2 are marked by their names. HC indicates healthy control; IS, ischemic stroke; qRT-PCR, quantitative real-time polymerase chain reaction; and TIA, transient ischemic attack.

were associated with stroke severity during hospitalization (day 1–7) as assessed by the National Institutes of Health stroke scale (miR-125b-5p: F=7.587, *P*=0.007; miR-143-3p: F=9.636, *P*=0.003; miR-125a-5p: F=0.855, *P*=0.358).

Relationship of Circulating MicroRNA Levels With Stroke Pathogenesis and Infarct Volume

To determine the relationship between circulating microRNAs and stroke pathogenesis, patients from the combined discovery and validation sample (n=60) and the replication sample (n=200) were stratified according to the Trial of Org 10172 in Acute Stroke Treatment classification system.[25] In both samples, levels of miR-125a-5p, miR-125b-5p, and miR-143-3p were similar across etiologic subgroups. This particularly applied to patients with large vessel stroke (n=61), cardioembolic stroke (n=79), and stroke of undetermined pathogenesis (n=96), whereas sample sizes for small-vessel stroke (n=18) and stroke of other determined

Table 2. Differentially Expressed Circulating MicroRNAs After Ischemic Stroke in the Discovery Sample

MicroRNA-ID	RNA Sequencing			qRT-PCR		Consistent Directionality of the FC
	Log_2 (CPM)	FC	FDR *P* Value	FC	*P* Value	
hsa-let-7b-3p	4.4	0.55	0.011	1.32	0.3107	…
hsa-let-7d-3p	7.8	0.60	0.014	1.83	4.9×10^{-6}*	No
hsa-let-7f-5p	11.6	0.42	<0.001	1.13	0.1767	…
hsa-let-7i-5p	10.2	0.48	0.005	1.58	0.0101	…
hsa-miR-1	4.4	0.16	<0.001	…	…	…
hsa-miR-16-2-3p	9.9	0.40	0.003	1.72	0.1798	…
hsa-miR-17-3p	3.0	0.40	0.011	…	…	…
hsa-miR-17-5p	5.0	0.37	0.003	1.77	0.0065	…
hsa-miR-18a-3p	3.5	0.38	0.004	…	…	…
hsa-miR-18a-5p	2.4	0.29	0.004	1.76	0.0508	…
hsa-miR-20a-5p	4.1	0.35	0.007	1.53	0.0635	…
hsa-miR-26b-5p	8.9	0.45	0.008	1.06	0.5800	…
hsa-miR-92a-3p	14.7	0.49	0.002	1.55	0.0453	…
hsa-miR-99b-5p	7.5	2.09	<0.001	1.65	0.0838	…
hsa-miR-101-3p	8.2	0.36	0.001	…	…	…
hsa-miR-125a-5p	7.8	2.47	<0.001	1.80	1.5×10^{-6}*	Yes
hsa-miR-125b-5p	5.0	1.69	0.026	2.54	5.6×10^{-6}*	Yes
hsa-miR-126-5p	10.1	0.46	<0.001	1.99	1.5×10^{-7}*	No
hsa-miR-130a-3p	7.5	0.57	0.022	3.73	0.0067	…
hsa-miR-140-3p	8.4	0.56	0.005	1.44	0.0678	…
hsa-miR-143-3p	10.8	1.88	0.049	4.77	7.8×10^{-9}*	Yes
hsa-miR-181a-5p	10.5	0.58	0.020	1.39	0.0113	…
hsa-miR-193a-5p	2.5	0.30	0.003	…	…	…
hsa-miR-378a-3p	6.8	0.58	0.026	1.23	0.0463	…
hsa-miR-423-3p	8.0	0.44	<0.001	2.14	4.2×10^{-6}*	No
hsa-miR-532-5p	5.4	0.33	<0.001	1.54	0.0119	…
hsa-miR-660-5p	5.0	0.29	<0.001	1.38	0.0582	…
hsa-miR-3158-3p	2.9	0.38	0.002	…	…	…
hsa-miR-3158-5p	2.9	0.38	0.002	…	…	…
hsa-miR-3184-5p	8.0	0.44	<0.001	…	…	…
hsa-miR-3688-3p	2.8	0.40	0.011	…	…	…
hsa-miR-3688-5p	2.8	0.40	0.011	…	…	…

Ten microRNAs were not detectable by qRT-PCR. Of the remaining 22 miRNAs, 6 (*) showed significant differential expression ($P<0.00227$). For 3 of them (miR-125a-5p, miR-125b-5p, and miR-143-3p), directionality of the FC was consistent between RNA sequencing and qRT-PCR. CPM indicates counts per million; FC, fold change; FDR, false discovery rate; and qRT-PCR, quantitative real-time polymerase chain reaction.

pathogenesis (n=6) were too small to draw conclusions (Online Figure VIII). Next, we determined whether levels of miR-125a-5p, miR-125b-5p, and miR-143-3p correlated with infarct volumes in the replication sample (n=188). Infarct volumes ranged from 0.045 to 385 mL. There was no significant correlation between the expression levels of individual microRNAs in the acute phase and infarct volumes (Online Figure IX).

Cellular Source and Extracellular Transportation of miR-125a-5p, miR-125b-5p, and miR-143-3p

In light of previous work suggesting platelets as an important source of circulating microRNAs,[26] we determined the extent to which platelets may contribute to circulating levels of miR-125a-5p, miR-125b-5p, and miR-143-3p. For this, we isolated platelets from platelet-rich plasma and spiked them back into platelet-poor plasma at increasing concentrations.[27]

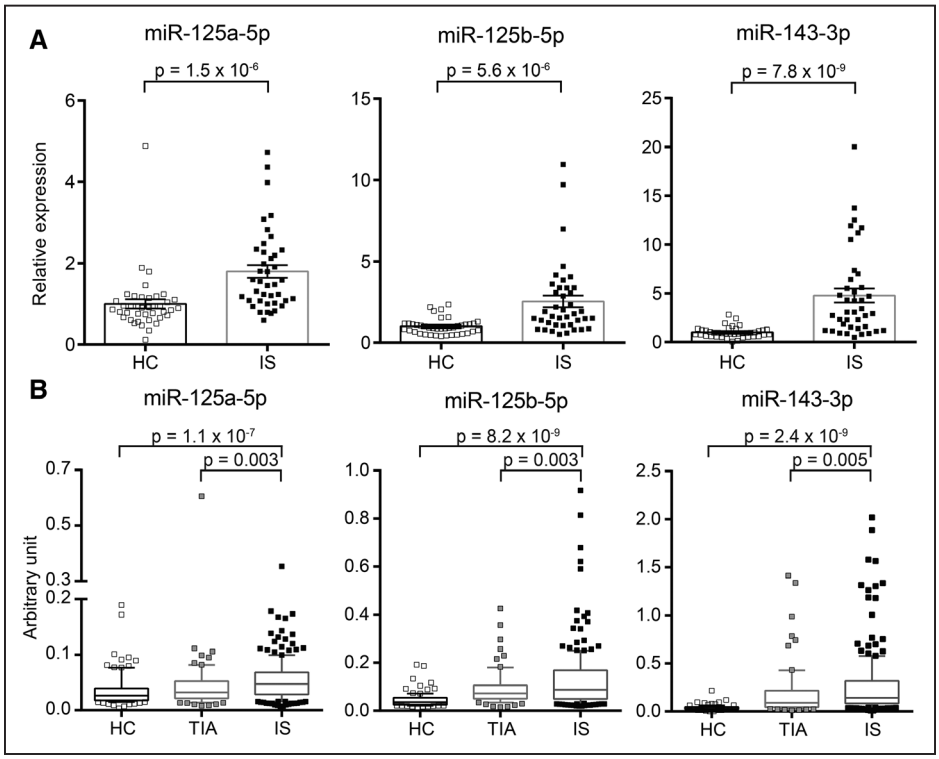


Figure 2. Validation and replication of miR-125a-5p, miR-125b-5p, and miR-143-3p. A, Results obtained in the independent validation sample (n=40/40, healthy control [HC]/ischemic stroke [IS]). miR-125a-5p, miR-125b-5p, and miR-143-3p all showed significant differential expression after stroke at hospital arrival ($P<0.00227$) with consistent directionality compared to RNA sequencing. **B**, Results obtained in the independent replication sample (n=100/72/200, HC/transient ischemic attack [TIA]/IS): miR-125a-5p, miR-125b-5p, and miR-143-3p. **A**, Mean±SEM, Mann–Whitney test; **B**, median±interquartile range, bars indicate 10th and 90th percentiles. Linear multivariate model after identification of covariates by backward stepwise regression.

Compared with liver-specific miR-122 (control), the levels of all 3 microRNAs significantly increased with increasing concentrations of platelets (Figure 4). To explore the possibility of a neuronal source of the 3 microRNAs during ischemia, we further challenged Neuro2a cells with increasing concentrations of the hypoxia-inducible factor 1-α stabilizing agent cobalt chloride for 2 and 6 hours and found no significant effect on the intra- and extracellular abundance of miR-125a-5p, miR-125b-5p, and miR-143-3p (Online Figure X). Also, we found no change of circulating levels of miR-125a-5p, miR-125b-5p, and miR-143-3p in 2 different experimental stroke models using transient proximal or permanent distal occlusion of the middle cerebral artery (Online Figure XI). These findings suggested platelets as a major source of the 3 microRNAs after acute IS. Next, to characterize the extracellular mode of transportation of miR-125a-5p, miR-125b-5p, and miR-143-3p, we isolated extracellular vesicles from patients with IS and HCs and found significantly higher levels of miR-143-3p in vesicles from patients with IS compared with vesicles from controls, whereas no significant difference was found for miR-125a-5p and miR-125b-5p (Online Figure XII).

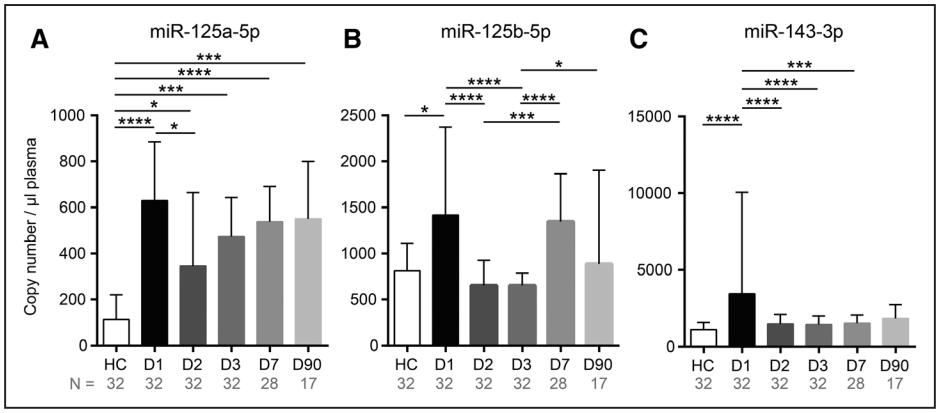


Figure 3. Temporal expression profile of validated microRNAs after stroke. Results for the 32 patients with ischemic stroke with available data for at least 4 time points after stroke: (**A**) miR-125a-5p, (**B**) miR-125b-5p, and (**C**) miR-143-3p. Median±interquartile range, Kruskal–Wallis test followed by Dunn multiple comparison test, *$P<0.05$, ***$P<0.001$, ****$P<0.0001$. D indicates day; and HC, healthy control.

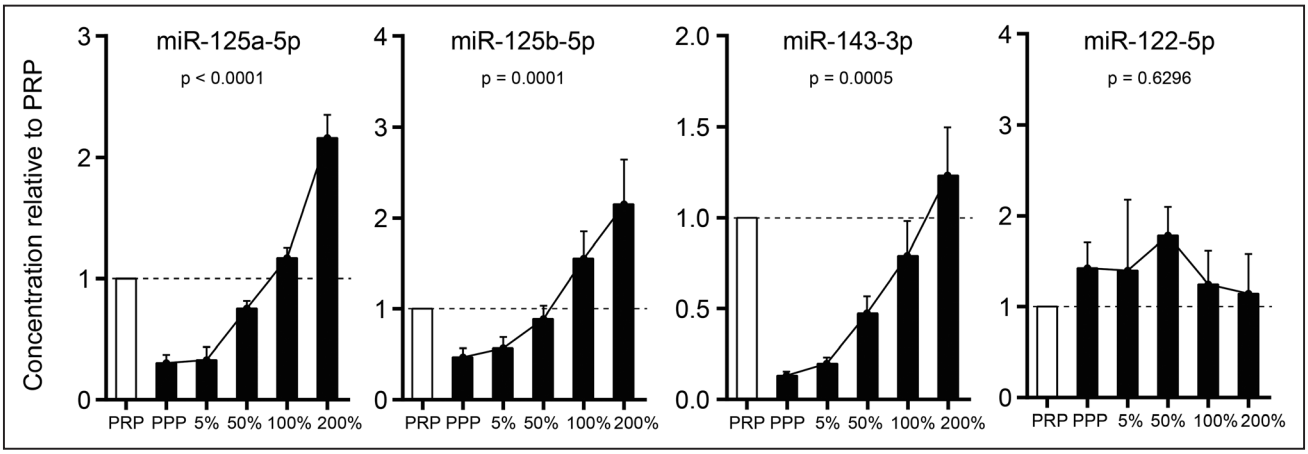


Figure 4. Platelet spike-in experiment. Levels of miR-125a-5p, miR-125b-5p, miR-143-3p, and miR-122-5p in relation to platelet-rich plasma (PRP) after spiking back platelets into platelet-poor plasma (PPP), n=3. Mean±SD, Friedman test.

Role of miR-125a-5p, miR-125b-5p, and miR-143-3p as a Diagnostic Marker for Acute IS

In view of elevated levels of circulating miR-125a-5p, miR-125b-5p, and miR-143-3p in the early phase after IS independent of stroke pathogenesis and infarct volume, we next determined their role as diagnostic markers in the acute setting by calculating the receiver operating characteristics curve. Using the discovery sample as a training set and the validation sample as a test set in a random forest classification, the signature of miR-125a-5p, miR-125b-5p, and miR-143-3p differentiated between HCs and patients with IS with an area under the curve (AUC) of 0.93 corresponding to a specificity of 97.5% and sensitivity of 65% (replication sample: AUC=0.90; sensitivity: 85.6%; specificity: 76.3%; Figure 5A and 5C). Adding the 3 microRNAs that were differentially expressed in the validation step but showed no consistent directionality with the discovery step resulted in no improvement of the receiver operating characteristics curve (Figure 5B). The AUC for differentiating between patients with TIA and IS was 0.66 (sensitivity: 89.2%; specificity: 27.1%; Figure 5D). For comparison, the combination of noncontrast CT, CT angiography, and CT perfusion obtained as part of clinical routine showed a sensitivity for IS of 55% (pooled discovery and validation sample) and 72.5% (replication sample), which were lower than obtained with the set of microRNAs (Online Table III). Of note, measurements of the previously reported biomarkers high-sensitivity C-reactive protein,[28,29] interleukin 6,[8] and neuron-specific enolase[5] yielded AUCs of 0.73, 0.82, and 0.69, respectively, which were lower than the AUC of the 3 microRNAs (Figure 5A and 5C). Adding interleukin 6 and neuron-specific enolase to the set of microRNAs further increased the AUC in the replication sample to 0.92 (sensitivity: 88.1%; specificity: 75.5%; Figure 5C) but did not improve the AUC for differentiating patients with IS from patients with TIA (AUC=0.66; Figure 5D).

Discussion

Our study establishes a set of circulating human microRNAs, which are differentially expressed after acute IS. Expression levels of miR-125a-5p, miR-125b-5p, and miR-143-3p were independent of infarct volume and stroke pathogenesis. When used to discriminate between patients with IS and HCs, this set showed a sensitivity of 85.6% and a specificity of 76.3% outperforming multimodal CT and previously reported biomarkers in our sample. We thus consider miR-125a-5p, miR-125b-5p, and miR-143-3p to have potential clinical utility as a diagnostic marker for acute IS.

To our knowledge, the receiver operating characteristics curve obtained for our set of microRNAs is unprecedented for a blood-based biomarker for stroke.[30,31] Previously reported blood biomarkers, such as neuron-specific enolase and glial fibrillary acidic protein, were shown to peak at ≥24 hours after IS with poor discrimination in the acute phase.[5,6] The potential of microRNAs as early diagnostic markers is illustrated by a recent study showing increased levels of specific microRNAs as early as 15 minutes after myocardial infarction.[32] Noncontrast CT is the imaging modality of choice for the initial assessment of patients with suspected stroke.[33] Multimodal CT has been shown to increase diagnostic performance in patients presenting with suspected IS.[2,34] However, its sensitivity in detecting signs of brain ischemia is only ≈60%.[2] In our sample, the combination of miR-125a-5p, miR-125b-5p, and miR-143-3p showed higher sensitivity than multimodal CT along with a high specificity.

Although the precise cellular sources and mechanisms underlying the observed elevations of miR-125a-5p, miR-125b-5p, and miR-143-3p remain to be determined, our study provides important clues. Specifically, we found that all 3 microRNAs depended on platelet numbers in a platelet spike-in experiment. Also, platelet counts and antiplatelet therapy were identified as relevant covariates in the microRNA expression analysis in our replication sample. In contrast, levels of miR-125a-5p, miR-125b-5p, and miR-143-3p were not affected by chemical hypoxia of Neuro2a cells in vitro and in 2 different experimental stroke models. Of note, experimental stroke models mimic some aspects of human IS, such as ischemic neuronal injury, whereas the primary mechanisms underlying vessel occlusion in humans, in particular thromboembolism, are not appropriately reflected by these models.[35] In fact, in the classical models used here, vessel occlusion was achieved by a filament (transient middle cerebral artery occlusion model) and thermocoagulation (permanent distal middle cerebral

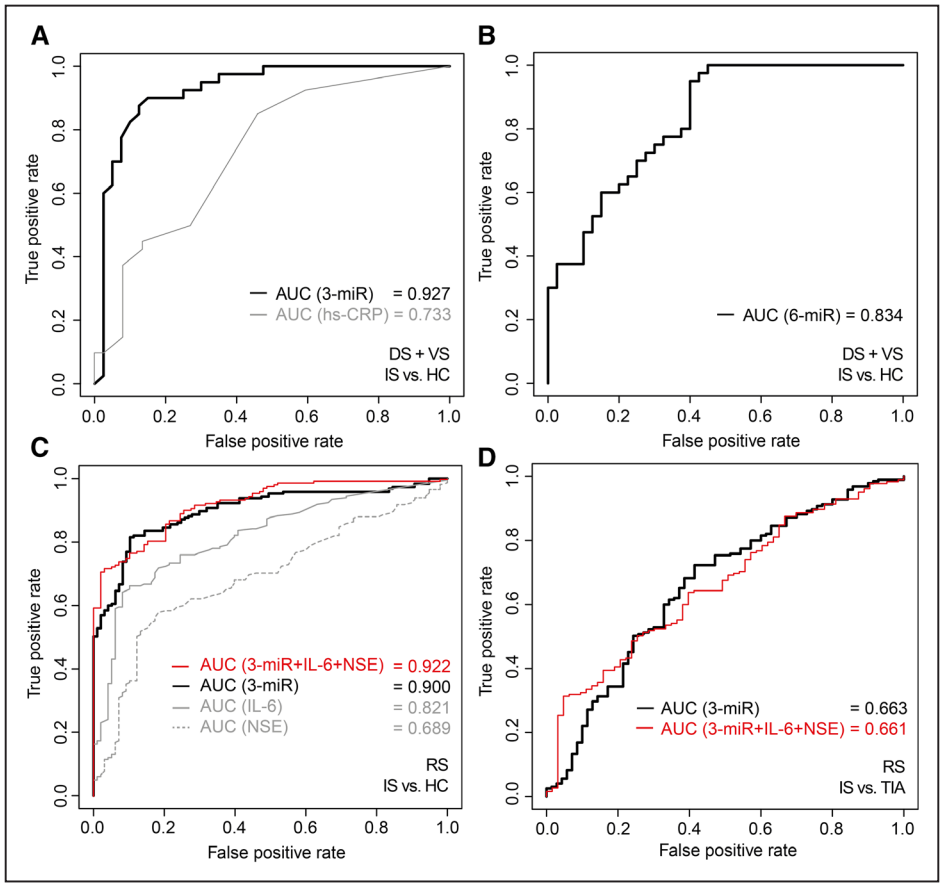


Figure 5. Diagnostic utility of a microRNA set for acute ischemic stroke (IS). Receiver operating characteristics were calculated using the expression levels of specific microRNAs at hospital arrival: (**A–C**) for differentiating patients with IS and healthy control subjects (HCs). **A**, miR-125a-5p, miR-125b-5p, and miR-143-3p using the discovery sample as training set and the validation sample as test set in comparison to high-sensitivity C-reactive protein (hs-CRP), (**B**) adding let-7d-3p, miR-126-5p, and miR-423-3p to the initial set because of their differential expression in the validation step but different directionality compared with the discovery step, (**C**) miR-125a-5p, miR-125b-5p, and miR-143-3p in the replication sample in comparison to and in combination with interleukin 6 (IL-6) and NSE (neuron-specific enolase), and (**D**) for differentiating patients with IS and patients with transient ischemic attack (TIA): miR-125a-5p, miR-125b-5p, and miR-143-3p compared with the combination of this set of miRNAs with IL-6 and NSE. 3-miR indicates the set of miR-125a-5p, miR-125b-5p, and miR-143-3p; 6-miR, set of miR-125a-5p, miR-125b-5p, miR-143-3p, let-7d-3p, miR-126-5p, and miR-423-3p; AUC, area under the curve; DS, discovery sample; RS, replication sample; and VS, validation sample.

artery occlusion model), thus not addressing the mechanisms of vessel occlusion in IS, which are difficult to model in mice. Collectively, our data suggest platelets as a major source of elevated miR-125a-5p, miR-125b-5p, and miR-143-3p in our patients. Whether this relates to platelet aggregation or thrombus formation remains to be determined.

Circulating microRNA levels could result from passive release or active secretion.[13] We found miR-143-3p levels to be significantly higher in extracellular vesicles of patients with IS compared with controls. Secreted microRNAs might serve as signaling molecules influencing systemic processes relevant for stroke outcome.[13] miR-143-3p has been linked to the phenotype and function of vascular smooth muscle cells[36] and endothelial cells.[37] Exosomes and nanotubes allow for shuttling of miR-143 between the 2 cell types leading to modulation of the angiogenic capacity of endothelial cells.[38,39] The miR-143/-145 cluster is integral for vascular smooth muscle cells to sustain a contractile phenotype.[40] Hence, it is tempting to speculate that elevated miR-143-3p levels may influence vessel function and structure after IS. miR-125a-5p and miR-125b-5p, displaying the same seed sequence, have both been shown to inhibit angiogenesis.[41,42] However, whereas miR-125a-5p has been linked to supporting endothelial barrier properties in the context of the blood–brain barrier[43] and to the differentiation of inflammatory cells,[44–46] miR-125b-5p regulates synaptic morphology and function.[47] Interestingly, miR-125a-5p remained elevated up to day 90, whereas expression levels of miR-125b-5p rapidly returned to control levels. Therefore, potential biological effects of miR-125a-5p might have more sustained relevance after acute stroke.

Importantly, we found the expression levels of miR-125a-5p, miR-125b-5p, and miR-143-3p to be independent of infarct volume and stroke pathogenesis. This finding emphasizes their potential utility as a broadly applicable diagnostic marker for acute IS. The observed absence of a correlation with infarct volume might seem counterintuitive but agrees with our experimental data, which showed that miR-125a-5p, miR-125b-5p, and miR-143-3p were not affected in paradigms of ischemic neuronal injury. Our findings further agree with an earlier observation showing that cardiac-specific miR-208a was elevated after myocardial infarction but did not correlate with the extent of myocardial injury.[32]

Patients with an unknown time of symptom onset (eg, wake-up stroke) pose a diagnostic challenge with implications for therapeutic decision making. Aside from showing a higher sensitivity in detecting IS on hospital arrival compared with CT, the signature of miR-125a-5p, miR-125b-5p, and miR-143-3p may provide information on the time interval since stroke onset. Specifically, elevated levels of all 3 microRNAs indicate the early phase after stroke and their peak of expression might be used to more accurately determine symptom onset and thus influence decisions on thrombolysis. However, this requires further investigation.

To our knowledge, the methodological rigor used in our study is unsurpassed in the stroke biomarker field: strengths include (1) the use of RNA sequencing in the discovery phase; (2) validation and replication in 2 independent samples; (3) inclusion of patients with TIA as the clinically most relevant differential diagnosis; (4) the exclusion of patients with prior antiplatelet therapy, cardiovascular events, or silent brain infarction in the discovery samples[48]; (5) consideration of other potential confounders of circulating microRNA levels; (6) serial sampling starting immediately on hospital admission before CT and medical interventions; and (7) direct comparison between RNA- and protein-based biomarkers.

Our study also has limitations. First, the quantification of microRNAs by RNA sequencing and quantitative real-time polymerase chain reaction is influenced by sample preparation, in particular RNA isolation[49] and cDNA preparation,[50] and dependent on the used platform.[51] However, by using 2 different methodologies, we aimed at controlling for these potential sources of bias and hence identifying true disease-related microRNAs. Second, subjects were recruited from a single center. Moreover, we cannot exclude that the expression levels of individual microRNAs had already been elevated before stroke onset. Of note, however, expression levels of miR-125b-5p and miR-143-3p rapidly returned to normal levels after the acute phase of stroke, thus evidencing a close relationship with the acute event.

In conclusion, circulating miR-125a-5p, miR-125b-5p, and miR-143-3p associated with acute IS with high specificity and with higher sensitivity than multimodal CT in our sample, thus identifying this set of microRNAs as a promising diagnostic marker.

Acknowledgments

We thank all members of the emergency department team at the Klinikum der Universität München, especially all nurses, for assistance with blood withdrawal and subsequent potential recruitment of patients with stroke. Also, we thank Uta Mamrak and Gemma Llovera for assistance with animal experiments.

Sources of Funding

M. Dichgans was supported by grants from the Deutsche Forschungsgemeinschaft (CRC 1123 [B3] and Munich Cluster for Systems Neurology [SyNergy]), the German Federal Ministry of Education and Research (BMBF, e:Med programme e:AtheroSysMed), the FP7/2007–2103 European Union project CVgenes@target (grant agreement No Health-F2-2013–601456), the European Union Horizon2020 projects SVDs@target (grant agreement no 66688) and CoSTREAM (grant agreement no 667375), the Fondation Leducq (Transatlantic Network of Excellence on the Pathogenesis of Small Vessel Disease of the Brain), the Vascular Dementia Research Foundation, and the Jackstaedt Foundation. S. Tiedt was supported by the Josef-Hackl-Stiftung.

Disclosures

None.

References

1. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388:1459–1544.
2. Hopyan J, Ciarallo A, Dowlatshahi D, Howard P, John V, Yeung R, Zhang L, Kim J, MacFarlane G, Lee TY, Aviv RI. Certainty of stroke diagnosis: incremental benefit with CT perfusion over noncontrast CT and CT angiography. *Radiology*. 2010;255:142–153. doi: 10.1148/radiol.09091021.
3. Wardlaw JM, Mielke O. Early signs of brain infarction at CT: observer reliability and outcome after thrombolytic treatment–systematic review. *Radiology*. 2005;235:444–453. doi: 10.1148/radiol.2352040262.
4. Whiteley W, Tseng MC, Sandercock P. Blood biomarkers in the diagnosis of ischemic stroke: a systematic review. *Stroke*. 2008;39:2902–2909. doi: 10.1161/STROKEAHA.107.511261.
5. Anand N, Stead LG. Neuron-specific enolase as a marker for acute ischemic stroke: a systematic review. *Cerebrovasc Dis*. 2005;20:213–219. doi: 10.1159/000087701.
6. Wunderlich MT, Wallesch CW, Goertler M. Release of glial fibrillary acidic protein is related to the neurovascular status in acute ischemic stroke. *Eur J Neurol*. 2006;13:1118–1123. doi: 10.1111/j.1468-1331.2006.01435.x.
7. Foerch C, Singer OC, Neumann-Haefelin T, du Mesnil de Rochemont R, Steinmetz H, Sitzer M. Evaluation of serum S100B as a surrogate marker for long-term outcome and infarct volume in acute middle cerebral artery infarction. *Arch Neurol*. 2005;62:1130–1134. doi: 10.1001/archneur.62.7.1130.
8. Smith CJ, Emsley HC, Gavin CM, Georgiou RF, Vail A, Barberan EM, del Zoppo GJ, Hallenbeck JM, Rothwell NJ, Hopkins SJ, Tyrrell PJ. Peak plasma interleukin-6 and other peripheral markers of inflammation in the first week of ischaemic stroke correlate with brain infarct volume, stroke severity and long-term outcome. *BMC Neurol*. 2004;4:2. doi: 10.1186/1471-2377-4-2.
9. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116:281–297.
10. Mitchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A*. 2008;105:10513–10518.
11. Zampetaki A, Willeit P, Drozdov I, Kiechl S, Mayr M. Profiling of circulating microRNAs: from single biomarkers to re-wired networks. *Cardiovasc Res*. 2012;93:555–562. doi: 10.1093/cvr/cvr266.
12. He Y, Lin J, Kong D, Huang M, Xu C, Kim TK, Etheridge A, Luo Y, Ding Y, Wang K. Current state of circulating microRNAs as cancer biomarkers. *Clin Chem*. 2015;61:1138–1155. doi: 10.1373/clinchem.2015.241190.
13. Creemers EE, Tijsen AJ, Pinto YM. Circulating microRNAs: novel biomarkers and extracellular communicators in cardiovascular disease? *Circ Res*. 2012;110:483–495. doi: 10.1161/CIRCRESAHA.111.247452.
14. Mayr M, Zampetaki A, Willeit P, Willeit J, Kiechl S. MicroRNAs within the continuum of postgenomics biomarker discovery. *Arterioscler Thromb Vasc Biol*. 2013;33:206–214. doi: 10.1161/ATVBAHA.112.300141.
15. Bhalala OG, Srikanth M, Kessler JA. The emerging roles of microRNAs in CNS injuries. *Nat Rev Neurol*. 2013;9:328–339. doi: 10.1038/nrneurol.2013.67.
16. Tan KS, Armugam A, Sepramaniam S, Lim KY, Setyowati KD, Wang CW, Jeyaseelan K. Expression profile of microRNAs in young stroke patients. *PLoS One*. 2009;4:e7689. doi: 10.1371/journal.pone.0007689.
17. Li P, Teng F, Gao F, Zhang M, Wu J, Zhang C. Identification of circulating microRNAs as potential biomarkers for detecting acute ischemic stroke. *Cell Mol Neurobiol*. 2015;35:433–447. doi: 10.1007/s10571-014-0139-5.
18. Tsai PC, Liao YC, Wang YS, Lin HF, Lin RT, Juo SH. Serum microRNA-21 and microRNA-221 as potential biomarkers for cerebrovascular disease. *J Vasc Res*. 2013;50:346–354. doi: 10.1159/000351767.
19. Li M, Zhang J. Circulating microRNAs: potential and emerging biomarkers for diagnosis of cardiovascular and cerebrovascular diseases. *Biomed Res Int*. 2015;2015:730535. doi: 10.1155/2015/730535.
20. Sacco RL, Kasner SE, Broderick JP, et al; American Heart Association Stroke Council, Council on Cardiovascular Surgery and Anesthesia; Council on Cardiovascular Radiology and Intervention; Council on

Cardiovascular and Stroke Nursing; Council on Epidemiology and Prevention; Council on Peripheral Vascular Disease; Council on Nutrition, Physical Activity and Metabolism. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2013;44:2064–2089. doi: 10.1161/STR.0b013e318296aeca.

21. Braun J, Misiak D, Busch B, Krohn K, Hüttelmaier S. Rapid identification of regulatory microRNAs by miTRAP (miRNA trapping by RNA in vitro affinity purification). *Nucleic Acids Res*. 2014;42:e66. doi: 10.1093/nar/gku127.
22. Stokowy T, Eszlinger M, Świerniak M, Fujarewicz K, Jarząb B, Paschke R, Krohn K. Analysis options for high-throughput sequencing in miRNA expression profiling. *BMC Res Notes*. 2014;7:144. doi: 10.1186/1756-0500-7-144.
23. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. doi: 10.1186/s13059-014-0550-8.
24. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–140. doi: 10.1093/bioinformatics/btp616.
25. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE 3rd. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24:35–41.
26. Zampetaki A, Willeit P, Tilling L, Drozdov I, Prokopi M, Renard JM, Mayr A, Weger S, Schett G, Shah A, Boulanger CM, Willeit J, Chowienczyk PJ, Kiechl S, Mayr M. Prospective study on circulating microRNAs and risk of myocardial infarction. *J Am Coll Cardiol*. 2012;60:290–299. doi: 10.1016/j.jacc.2012.03.056.
27. Kaudewitz D, Skroblin P, Bender LH, et al. Association of microRNAs and YRNAs with platelet function. *Circ Res*. 2016;118:420–432. doi: 10.1161/CIRCRESAHA.114.305663.
28. Worthmann H, Tryc AB, Goldbecker A, Ma YT, Tountopoulou A, Hahn A, Dengler R, Lichtinghagen R, Weissenborn K. The temporal profile of inflammatory markers and mediators in blood after acute ischemic stroke differs depending on stroke outcome. *Cerebrovasc Dis*. 2010;30:85–92. doi: 10.1159/000314624.
29. Glickman SW, Phillips S, Anstrom KJ, Laskowitz DT, Cairns CB. Discriminative capacity of biomarkers for acute stroke in the emergency department. *J Emerg Med*. 2011;41:333–339. doi: 10.1016/j.jemermed.2010.02.025.
30. Dvorak F, Haberer I, Sitzer M, Foerch C. Characterisation of the diagnostic window of serum glial fibrillary acidic protein for the differentiation of intracerebral haemorrhage and ischaemic stroke. *Cerebrovasc Dis*. 2009;27:37–41. doi: 10.1159/000172632.
31. González-García S, González-Quevedo A, Peña-Sánchez M, Menéndez-Saínz C, Fernández-Carriera R, Arteche-Prior M, Pando-Cabrera A, Fernández-Concepción O. Serum neuron-specific enolase and S100 calcium binding protein B biomarker levels do not improve diagnosis of acute stroke. *J R Coll Physicians Edinb*. 2012;42:199–204. doi: 10.4997/JRCPE.2012.302.
32. Liebetrau C, Möllmann H, Dörr O, Szardien S, Troidl C, Willmer M, Voss S, Gaede L, Rixe J, Rolf A, Hamm C, Nef H. Release kinetics of circulating muscle-enriched microRNAs in patients undergoing transcoronary ablation of septal hypertrophy. *J Am Coll Cardiol*. 2013;62:992–998. doi: 10.1016/j.jacc.2013.05.025.
33. Jauch EC, Saver JL, Adams HP Jr, Bruno A, Connors JJ, Demaerschalk BM, Khatri P, McMullan PW Jr, Qureshi AI, Rosenfield K, Scott PA, Summers DR, Wang DZ, Wintermark M, Yonas H; American Heart Association Stroke Council; Council on Cardiovascular Nursing; Council on Peripheral Vascular Disease; Council on Clinical Cardiology. 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*. 2013;44:870–947. doi: 10.1161/STR.0b013e318284056a.
34. Zhu G, Michel P, Aghaebrahim A, Patrie JT, Xin W, Eskandari A, Zhang W, Wintermark M. Computed tomography workup of patients suspected of acute ischemic stroke: perfusion computed tomography adds value compared with clinical evaluation, noncontrast computed tomography, and computed tomography angiogram in terms of predicting outcome. *Stroke*. 2013;44:1049–1055. doi: 10.1161/STROKEAHA.111.674705.
35. Sommer CJ. Ischemic stroke: experimental models and reality. *Acta Neuropathol*. 2017;133:245–261. doi: 10.1007/s00401-017-1667-0.
36. Zampetaki A, Mayr M. MicroRNAs in vascular and metabolic disease. *Circ Res*. 2012;110:508–522. doi: 10.1161/CIRCRESAHA.111.247445.
37. Ramanujam D, Engelhardt S. Intercellular miRNA traffic. *Circ Res*. 2015;116:1726–1728. doi: 10.1161/CIRCRESAHA.115.306519.
38. Deng L, Blanco FJ, Stevens H, et al. MicroRNA-143 activation regulates smooth muscle and endothelial cell crosstalk in pulmonary arterial hypertension. *Circ Res*. 2015;117:870–883. doi: 10.1161/CIRCRESAHA.115.306806.
39. Climent M, Quintavalle M, Miragoli M, Chen J, Condorelli G, Elia L. TGFβ triggers miR-143/145 transfer from smooth muscle cells to endothelial cells, thereby modulating vessel stabilization. *Circ Res*. 2015;116:1753–1764. doi: 10.1161/CIRCRESAHA.116.305178.
40. Cordes KR, Sheehy NT, White MP, Berry EC, Morton SU, Muth AN, Lee TH, Miano JM, Ivey KN, Srivastava D. miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature*. 2009;460:705–710. doi: 10.1038/nature08195.
41. Che P, Liu J, Shan Z, Wu R, Yao C, Cui J, Zhu X, Wang J, Burnett MS, Wang S, Wang J. miR-125a-5p impairs endothelial cell angiogenesis in aging mice via RTEF-1 downregulation. *Aging Cell*. 2014;13:926–934. doi: 10.1111/acel.12252.
42. Muramatsu F, Kidoya H, Naito H, Sakimoto S, Takakura N. MicroRNA-125b inhibits tube formation of blood vessels through translational suppression of VE-cadherin. *Oncogene*. 2013;32:414–421. doi: 10.1038/onc.2012.68.
43. Reijerkerk A, Lopez-Ramirez MA, van Het Hof B, Drexhage JA, Kamphuis WW, Kooij G, Vos JB, van der Pouw Kraan TC, van Zonneveld AJ, Horrevoets AJ, Prat A, Romero IA, de Vries HE. MicroRNAs regulate human brain endothelial cell-barrier function in inflammation: implications for multiple sclerosis. *J Neurosci*. 2013;33:6857–6863. doi: 10.1523/JNEUROSCI.3965-12.2013.
44. Pan W, Zhu S, Dai D, et al. MiR-125a targets effector programs to stabilize Treg-mediated immune homeostasis. *Nat Commun*. 2015;6:7096. doi: 10.1038/ncomms8096.
45. Banerjee S, Cui H, Xie N, Tan Z, Yang S, Icyuz M, Thannickal VJ, Abraham E, Liu G. miR-125a-5p regulates differential activation of macrophages and inflammation. *J Biol Chem*. 2013;288:35428–35436. doi: 10.1074/jbc.M112.426866.
46. Graff JW, Dickson AM, Clay G, McCaffrey AP, Wilson ME. Identifying functional microRNAs in macrophages with polarized phenotypes. *J Biol Chem*. 2012;287:21816–21825. doi: 10.1074/jbc.M111.327031.
47. Edbauer D, Neilson JR, Foster KA, Wang CF, Seeburg DP, Batterton MN, Tada T, Dolan BM, Sharp PA, Sheng M. Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. *Neuron*. 2010;65:373–384. doi: 10.1016/j.neuron.2010.01.005.
48. Willeit P, Zampetaki A, Dudek K, et al. Circulating microRNAs as novel biomarkers for platelet activation. *Circ Res*. 2013;112:595–600. doi: 10.1161/CIRCRESAHA.111.300539.
49. McAlexander MA, Phillips MJ, Witwer KW. Comparison of methods for miRNA extraction from plasma and quantitative recovery of RNA from cerebrospinal fluid. *Front Genet*. 2013;4:83. doi: 10.3389/fgene.2013.00083.
50. Hafner M, Renwick N, Brown M, Mihailović A, Holoch D, Lin C, Pena JT, Nusbaum JD, Morozov P, Ludwig J, Ojo T, Luo S, Schroth G, Tuschl T. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA*. 2011;17:1697–1712. doi: 10.1261/rna.2799511.
51. Mestdagh P, Hartmann N, Baeriswyl L, et al. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat Methods*. 2014;11:809–815. doi: 10.1038/nmeth.3014.

Circulation Research
JOURNAL OF THE AMERICAN HEART ASSOCIATION

RNA-Seq Identifies Circulating miR-125a-5p, miR-125b-5p, and miR-143-3p as Potential Biomarkers for Acute Ischemic Stroke

Steffen Tiedt, Matthias Prestel, Rainer Malik, Nicola Schieferdecker, Marco Duering, Veronika Kautzky, Ivelina Stoycheva, Julia Böck, Bernd H. Northoff, Matthias Klein, Franziska Dorn, Knut Krohn, Daniel Teupser, Arthur Liesz, Nikolaus Plesnila, Lesca Miriam Holdt and Martin Dichgans

Circ Res. 2017;121:970-980; originally published online July 19, 2017;
doi: 10.1161/CIRCRESAHA.117.311572
Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0009-7330. Online ISSN: 1524-4571

SUPPLEMENTAL MATERIAL

SUPPLEMENTAL METHODS

Study population

Ischemic stroke (IS) patients, transient ischemic attack (TIA) patients and healthy control subjects (HCs) were recruited at the Klinikum der Universität München (KUM), a tertiary level hospital at Ludwig-Maximilians-Universität, Munich, Germany. The enrolment period was January 2014 to January 2017. For the discovery and validation sample, we excluded IS patients and HCs with active malignant disease, inflammatory or infectious diseases, surgery within the last three months and prior medication with low-molecular or unfractionated heparin within the last month. For the discovery sample we further excluded IS patients and HCs with prior use of antiplatelet medication within the last month, a history of myocardial infarction, stroke, or transient ischemic attack, or signs for silent CNS infarction on neuroimaging. For the replication sample, we only excluded patients with prior medication with low-molecular or unfractionated heparin within the last month. No study subject suffered from polycythemia vera or essential thrombocythemia. The study was approved by the local ethics committee and was conducted in accordance with the Declaration of Helsinki as well as institutional guidelines. Written and informed consent was obtained from all subjects.

Clinical assessments

All IS patients received a comprehensive diagnostic work-up (magnetic resonance imaging: 62.3 %, non-contrast-enhanced computed tomography (CT): 98.5 %, CT angiography: 96.2 %, CT perfusion: 74.2 %, 24h-electrocardiography: 31.5 %, transthoracic or transesophageal echocardiography: 57.3 %, extra- and transcranial ultrasound: 81.2 %). Stroke etiology was classified according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST).[1]

Patient blood sampling and processing

Blood samples from IS and TIA patients were collected at hospital arrival in the emergency department. Longitudinal sampling of IS patients was additionally performed at the second, third and seventh day of hospitalization (or at earlier discharge) and at 90 days post-stroke. Samples from HCs were taken in the outpatient clinic of KUM. Whole blood was drawn into EDTA-plasma containers (Sarstedt) using a tourniquet and 21 gauge needles. After 30 to 45 minutes at room temperature separation of plasma was achieved by differential centrifugation at 2000*g* for 10 minutes and 2500*g* for 15 minutes at 15 °C. Samples were aliquoted in screw cap vials and kept at -80 °C. Serum sampling was achieved by blood collection in serum containers (Sarstedt) and subsequent centrifugation at 2000*g* for 10 minutes.

Isolation of extracellular vesicles

Extracellular vesicles from plasma were isolated using the miRCURY Exosome Isolation kit – Serum and Plasma (Exiqon) following the manufacturer's instructions. In brief, from both groups, IS patients and HCs, 5 pools each consisting of 6 individual plasma samples from the validation sample, were used. Samples were incubated with a precipitation buffer at 4 °C for at least 60 minutes and centrifuged for 5 minutes at 500g. The pelleted vesicles were resuspended using a resuspension buffer.

Platelet spike-in experiment

The experiment was adopted from a previously published protocol.[2] In short, whole blood from 3 healthy volunteers was drawn into EDTA-plasma containers (Sarstedt) using a tourniquet and 21 gauge needles. After 30 minutes at room temperature, samples were centrifuged at 190g for 30 minutes at room temperature to pellet red blood cells. Next, the supernatant was centrifuged at 280g for 10 minutes to deplete white blood cells. The volume of the supernatant, consisting of platelet-rich plasma (PRP), was measured and 200 µl were kept for RNA extraction. The remainder was centrifuged at 1180g for 10 minutes to pellet platelets. The obtained supernatant was platelet-poor plasma (PPP). To further purify platelets, the pellet was washed with PBS and sorted by CD61-labeled beads (Miltenyi Biotec, #130-051-101) according to the manufacturer's protocol (Miltenyi Biotec). The final platelet solution was again pelleted and resuspended in 1/20 of the initial PRP volume.

Eventually, this stock solution was used to reconstitute PPP with platelet concentrations of 5 %, 50 %, 100 %, and 200 %. Subsequently, RNA was extracted as described for plasma.

Animal experiments
All animal experiments were approved by the governmental committee of Upper Bavaria (protocol number 55.2.1.54-2532-20-14) and were performed in accordance with guidelines for the use of experimental animals. Wild-type C57BL/6 mice were obtained from Charles River Laboratories. Animals were randomized to surgery groups and subsequent experiments were performed by an investigator blinded with respect to the surgery group. All animal experiments were reported consistent with the ARRIVE guidelines (Kilkenny et al., 2010).

Transient fMCAO
Mice were anesthetized with isoflurane delivered in a mixture of 30 % O_2 and 70 % N_2O. An incision was made between the ear and the eye to expose the temporal bone. A laser Doppler probe was affixed to the skull above the middle cerebral artery (MCA) territory and the mice were placed in the supine position. An incision was made in the midline neck region and the common carotid artery and left external carotid artery were isolated and ligated; a 2 mm silicon-coated filament (Doccol, #701912PKRe) was inserted into the internal carotid artery and MCA occlusion was confirmed by a corresponding decrease in blood flow (i.e., a decrease in the laser Doppler flow signal to less than 20 % of baseline). After 60 minutes of occlusion, the animals were reanesthetized and the filament was removed. For the survival period, the mice were kept in their home cage with facilitated access to water and food. Sham-operated mice received the same surgical procedure except the filament was inserted and immediately removed. Body temperature was maintained at 37 °C throughout surgery using a feedback-controlled heating pad. The overall mortality rate in mice subjected to MCAo was 10 %. Exclusion criteria were as follows: insufficient MCA occlusion (a reduction in cerebral blood flow to > 20% of the baseline value; n=0) and mortality during surgery (n=0).

Permanent distal MCA occlusion model
Focal cerebral ischemia was induced as described previously by permanent occlusion of the MCA distal of the lenticulostriate arteries (Llovera et al., 2014). In brief, the mice were anesthetized with isoflurane delivered in a mixture of 30 % O_2 and 70 % N_2O. The skull was exposed by a lateral skin incision between eye and ear, a burr hole was drilled in the temporal bone, and the MCA was permanently occluded using high-frequency electrocoagulation forceps (ERBE Erbotome). After recovery, mice were returned to their home cages with ad libitum access to water and food. Sham surgery was performed by the same surgical procedures without coagulation of the exposed MCA. Throughout the surgical procedure, body temperature was maintained at 37 °C using a feedback-controlled heating pad. The overall mortality rate in this experimental group was 5 %. Exclusion criteria were subarachnoid hemorrhage or death during surgery (n=1).

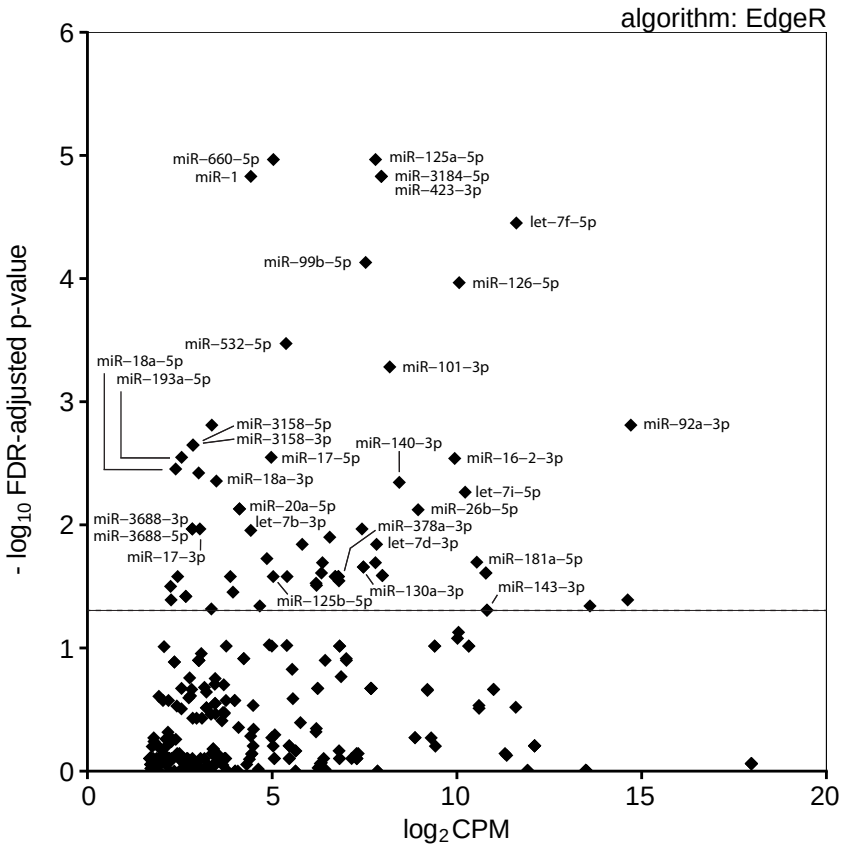
Animal blood sampling and processing
Under deep anethesia with ketamine (120 mg/kg) and xylazine (16 mg/kg), whole blood from mice was collected by cardiac puncture of the right ventricle using 25 gauge needles and drawn into EDTA-plasma containers (Sarstedt). After 30 to 45 minutes at room temperature separation of plasma was achieved by differential centrifugation at 2000*g* for 10 minutes and 2500*g* for 15 minutes at 15 °C. Samples were kept at -80 °C.

RNA isolation from plasma and reverse transcription
RNA < 1000 nucleotides was isolated from 200 µl plasma using the miRCURY RNA Isolation Kit – Biofluids (Exiqon) according to the manufacturer's recommendations with additional use of 3 µg glycogen as carrier.[3] The process was monitored by using the manufacturer's spike-in mix (UniSp2, UniSp4 and UniSp5). Starting from day 2 most IS patients received low-molecular-weight heparin. Hence, all samples included into the longitudinal analyses were treated with heparinase for one hour at room temperature prior to RNA isolation.[4,5] The reverse transcription reaction was performed using the Universal cDNA Synthesis Kit II (Exiqon) following the manufacturer's instructions.

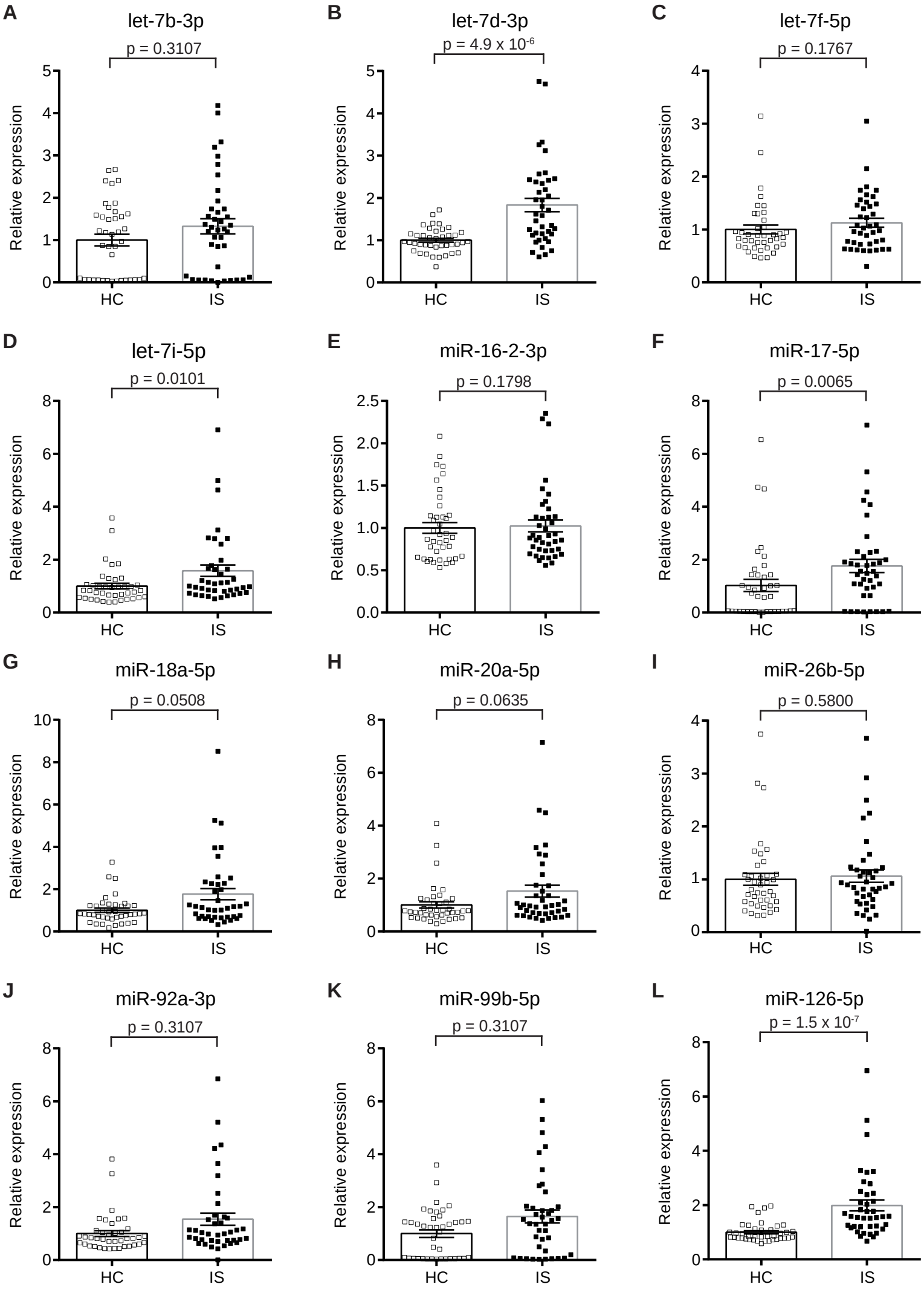
qPCR assay
ExiLENT SYBR® Green and microRNA LNA™ PCR primer sets (Exiqon) were used to determine microRNA (miRNA) expression levels in a LightCycler 480/II (Roche). All samples were run in duplicates. The spiked-in UniSp2, UniSp4 and UniSp5 were used for normalization. Cq-values from miRNAs of interest were normalized using a median normalization method (discovery and validation step) or the ΔΔCq normalization method (replication step). Replicates of individual samples with Cq values > 35 were set to 40. miRNAs were excluded if more than half of the samples within each group showed Cq values > 35. miRNA copy numbers were calculated using recombinant miRNAs (Metabion) as standards.

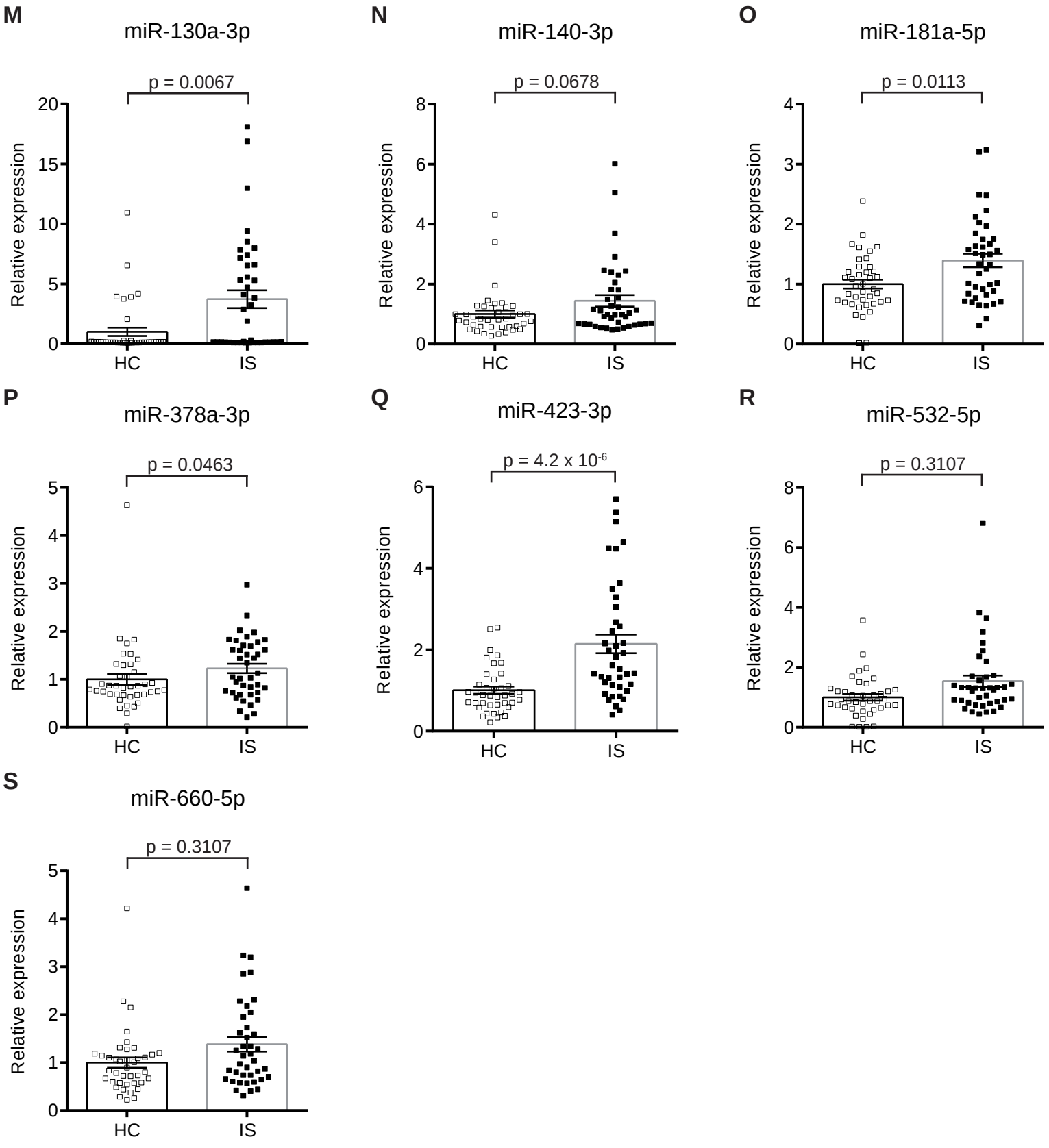
Quantification of infarct volumes
Infarct volumes were quantified on images from diagnostic scans, either CT or magnetic resonance imaging (diffusion-weighted, T2 or fluid-attenuated inversion recovery). The modality and image with the largest infarct size was used for volumetry. Trained raters segmented infarcts manually slice-by-slice. The inter-rater reliability for this procedure showed an intraclass correlation coefficient of 0.993.

Statistics
Data for continuous variables are expressed as mean +/- standard error of the mean or median +/- interquartile range, if not stated otherwise. For group-wise comparisons (n > 2 groups) ANOVA or Kruskal–Wallis test followed by Dunn's multiple comparison test were used as appropriate. For n = 2 groups Mann–Whitney test or Student's t test were used as appropriate. Outliers were identified by the ROUT method. Matched data were analyzed using the Durbin-Skillings-Mack-test or Friedman test. For categorical variables, we used the Fisher's exact test or the Chi-squared test. All tests were performed 2-sided. A p-value < 0.05 was considered statistically significant. If indicated, Bonferroni correction for multiple testing was applied. Random forest classification (R package 'randomForest')[6] was used for constructing receiver operating characteristic statistics. We used standard parameters and grew 100,001 trees per run. Linear and polynomial regressions as well as generalized linear mixed models were used to characterize the relationship of miRNA expression levels with infarct volumes, time since symptom onset and stroke severity (as measured by the NIH stroke scale (NIHSS)). For multivariate linear regression, we used a backward stepwise approach in order to identify covariates building the model with the best fit as indicated by the Akaike information criterion (AIC) ('StepAIC' function from the 'MASS' package in 'R').[7] Statistical analyses were done using GraphPad Prism 6.0 for Windows, SPSS Statistics 24 or the programming environment 'R', version 3.4.0.[8]

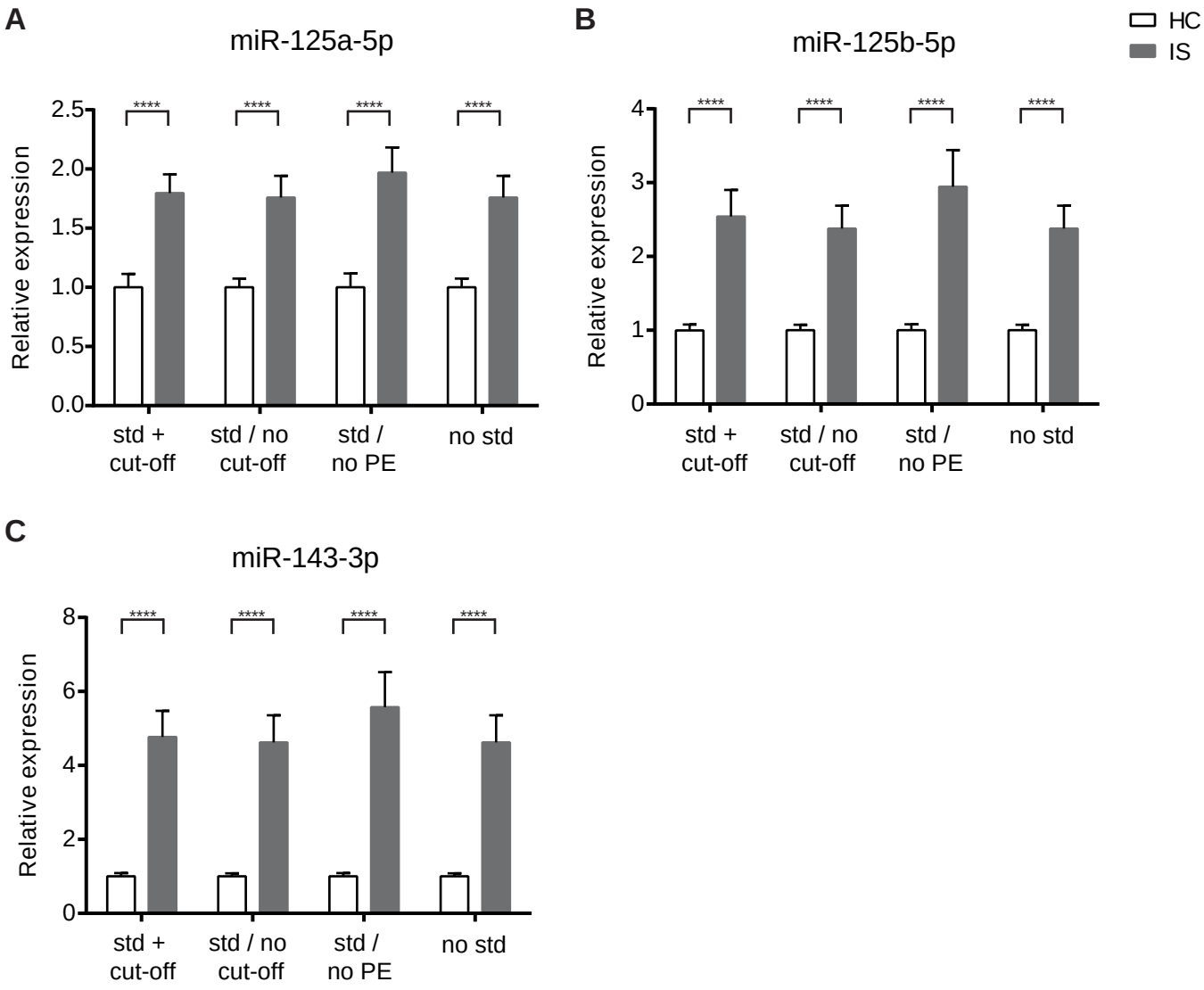


Online Figure I. Scatter plot of RNA sequencing results. Individual miRNAs are displayed by their FDR-adjusted p-value (for the comparison of ischemic stroke patients vs. healthy controls) and the corresponding CPM (normalization algorithm EdgeR). The dashed line indicates an FDR-adjusted p-value of 0.05. miRNAs that showed differential expression in both EdgeR and DESeq2 are marked by their names. miRNA, microRNA; FDR, false discovery rate; CPM, counts per million.

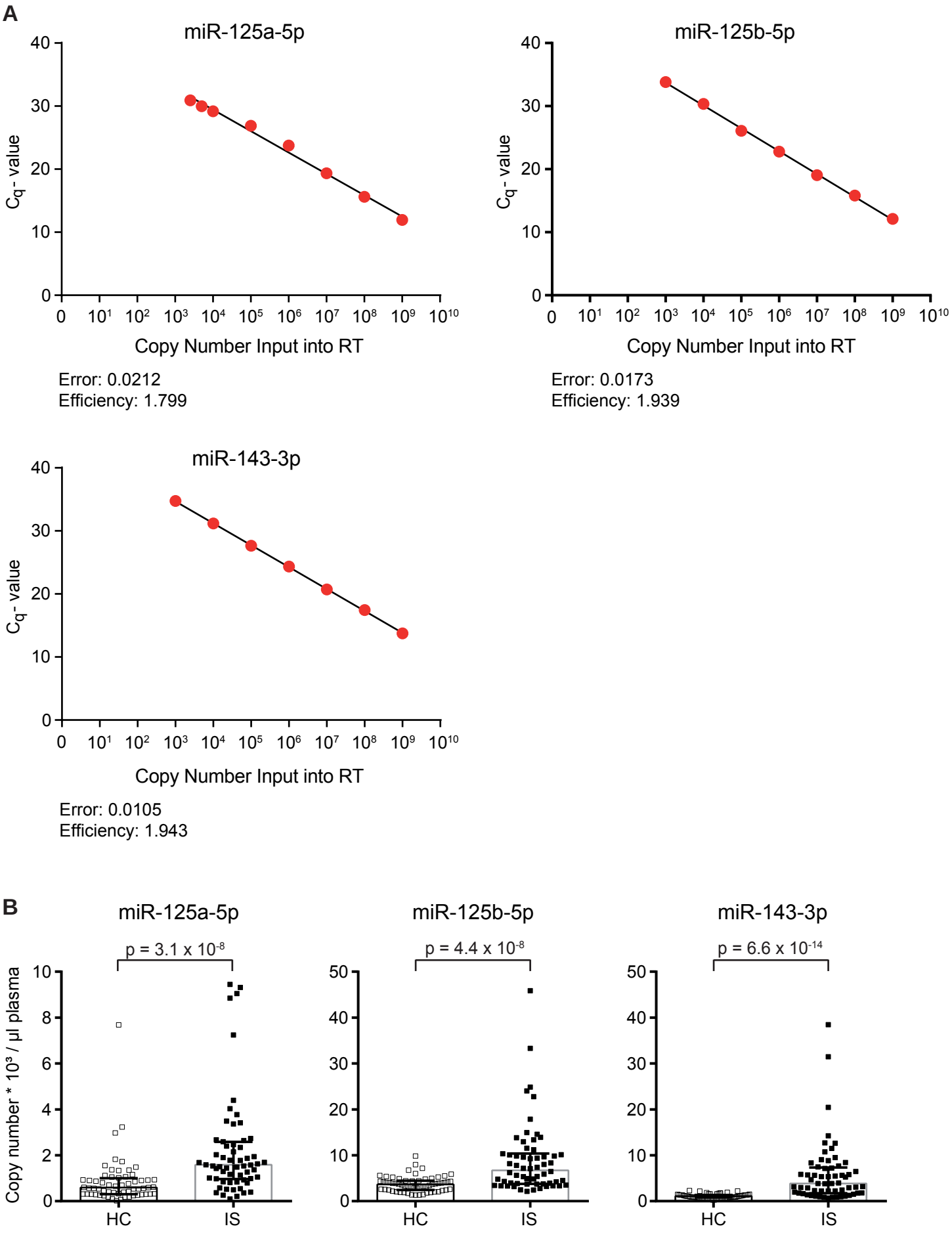

A
let-7b-3p
p = 0.3107
B
let-7d-3p
p = 4.9 x 10-6
C
let-7f-5p
p = 0.1767
D
let-7i-5p
p = 0.0101
E
miR-16-2-3p
p = 0.1798
F
miR-17-5p
p = 0.0065
G
miR-18a-5p
p = 0.0508
H
miR-20a-5p
p = 0.0635
I
miR-26b-5p
p = 0.5800
J
miR-92a-3p
p = 0.3107
K
miR-99b-5p
p = 0.3107
L
miR-126-5p
p = 1.5 x 10-7
Relative expression
HC
IS



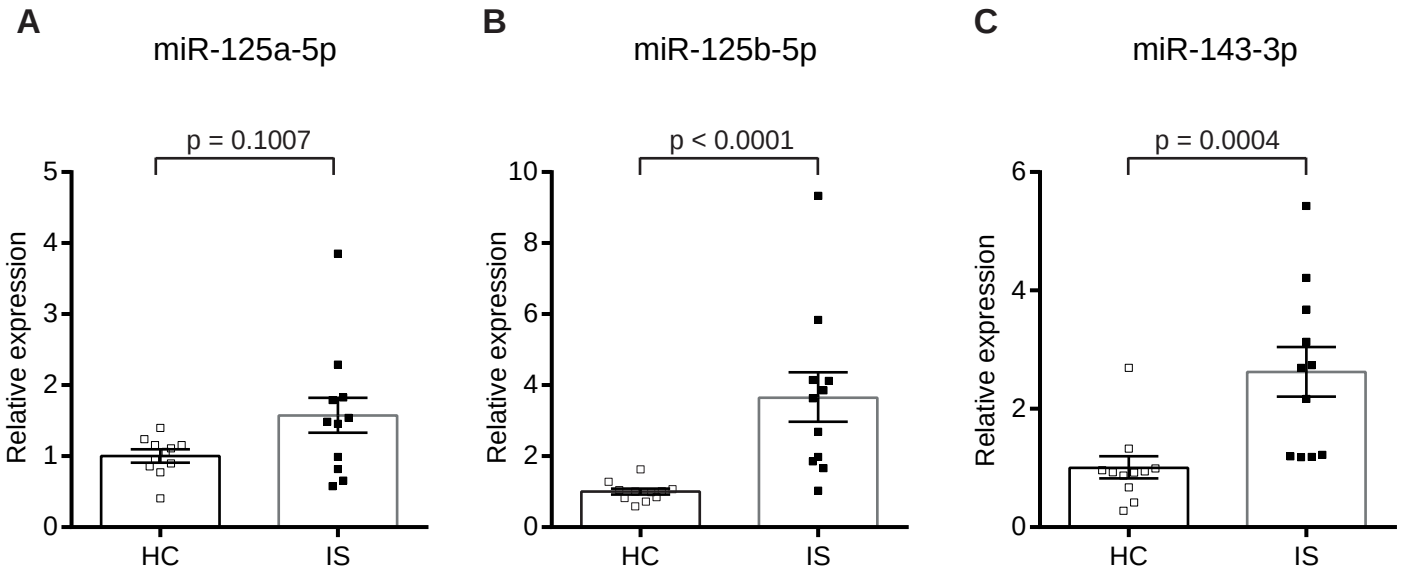
Online Figure II. Validation experiment results of miRNAs that were selected via RNA sequencing but not confirmed via qRT-PCR. 16 miRNAs were not significantly differentially expressed: (**A**) let-7b-3p, (**C**) let-7f-5p, (**D**) let-7i-5p, (**E**) miR-16-2-3p, (**F**) miR-17-5p, (**G**) miR-18a-5p, (**H**) miR-20a-5p, (**I**) miR-26b-5p, (**J**) miR-92a-3p, (**K**) miR-99b-5p, (**M**) miR-130a-5p, (**N**) miR-140-3p, (**O**) miR-181a-5p, (**P**) miR-378a-3p, (**R**) miR-532-3p, and (**S**) miR-660-5p. Three circulating miRNAs were significantly differentially expressed ($p < 0.00227$) but did not show consistent directionality with RNA sequencing: (**B**) let-7d-3p, (**L**) miR-126-5p, and (**Q**) miR-423-3p. Mean ± SEM, N=40, Mann-Whitney test. miRNA, microRNA; qRT-PCR, quantitative real-time polymerase chain reaction; SEM, standard error of the mean; groups are indicated by: IS, ischemic stroke; and HC, healthy control subjects.



Online Figure III. Alternative analyses for the validation of miR-125a-5p, miR-125b-5p and miR-143-3p. The validation results for (**A**) miR-125a-5p, (**B**) miR-125b-5p, and (**C**) miR-143-3p remained stable (compared to the established analysis (std + cut-off)) when not correcting qRT-PCR data with Cq values > 35 (std / no cut-off), when excluding subjects with previous major cardiovascular events (std / no PE) and when using non-standardized qRT-PCR data (without correction by Spike-Ins; no std). Mean ± SEM, N=40, Mann-Whitney test, **** $p < 0.0001$. qRT-PCR, quantitative real-time polymerase chain reaction; SEM, standard error of the mean; groups are indicated by: IS, ischemic stroke; and HC, healthy control subjects. Std, standardized by correction with spiked-in templates; PE, previous major cardiovascular events.



Online Figure IV. Standard curves and qRT-PCR validation for miR-125a-5p, miR-125b-5p and miR-143-3p. (**A**) Different copy number inputs of recombinant miRNAs were used to generate standard curves for miR-125a-5p, miR-125b-5p, and miR-143-3p. (**B**) Copy numbers per microliter plasma obtained in the combined discovery and validation samples (N=60): miR-125a-5p, miR-125b-5p, and miR-143-3p. Median ± IQR; Mann-Whitney test. qRT-PCR, quantitative real-time polymerase chain reaction; RT, reverse transcription reaction; miRNA, microRNA; IQR, interquartile range; groups are indicated by: IS, ischemic stroke; and HC, healthy control subjects.

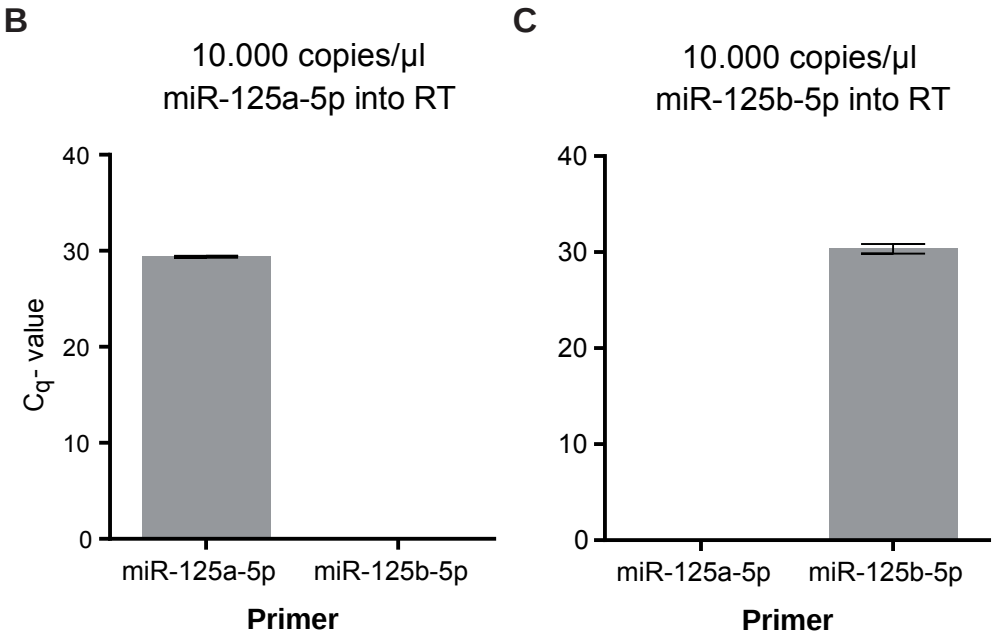


Online Figure V. Serum expression levels of miR-125a-5p, miR-125b-5p and miR-143-3p. To exclude plasma-specific differential miRNA expression, 10 IS patients and HCs were randomly selected from the validation sample and their serum samples analyzed for the expression levels of (**A**) miR-125a-5p, (**B**) miR-125b-5p, and (**C**) miR-143-3p. Mean ± SEM, N=10, Mann-Whitney test. miRNA, microRNA; SEM, standard error of the mean; IS, ischemic stroke; HC, healthy control subjects.

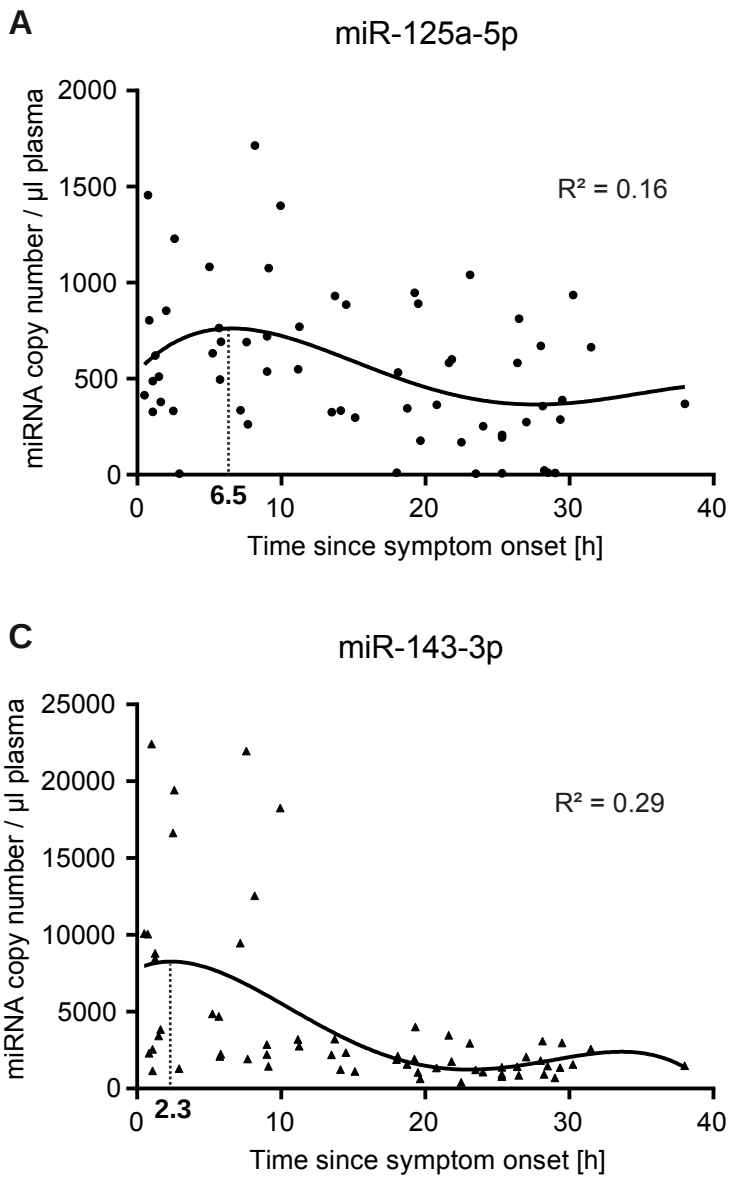
A

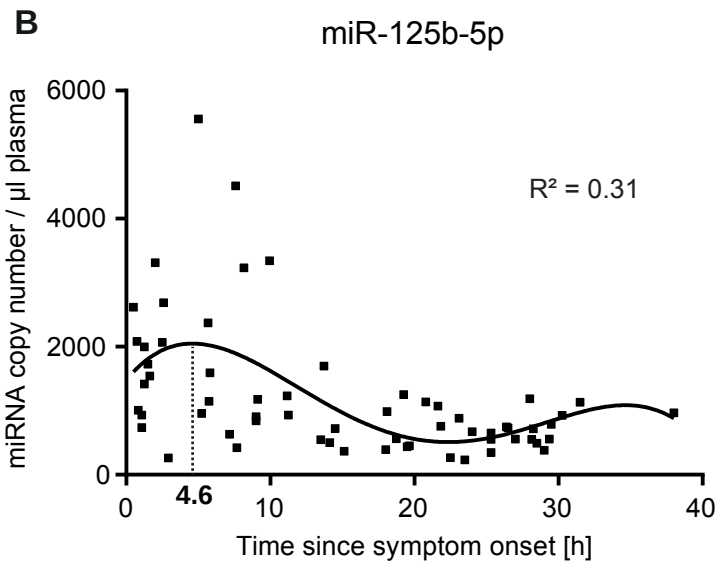
hsa-miR-125a-5p: ucccugagacccuuuaaccuguga

hsa-miR-125b-5p: ucccugagacccu_ _aacuuguga

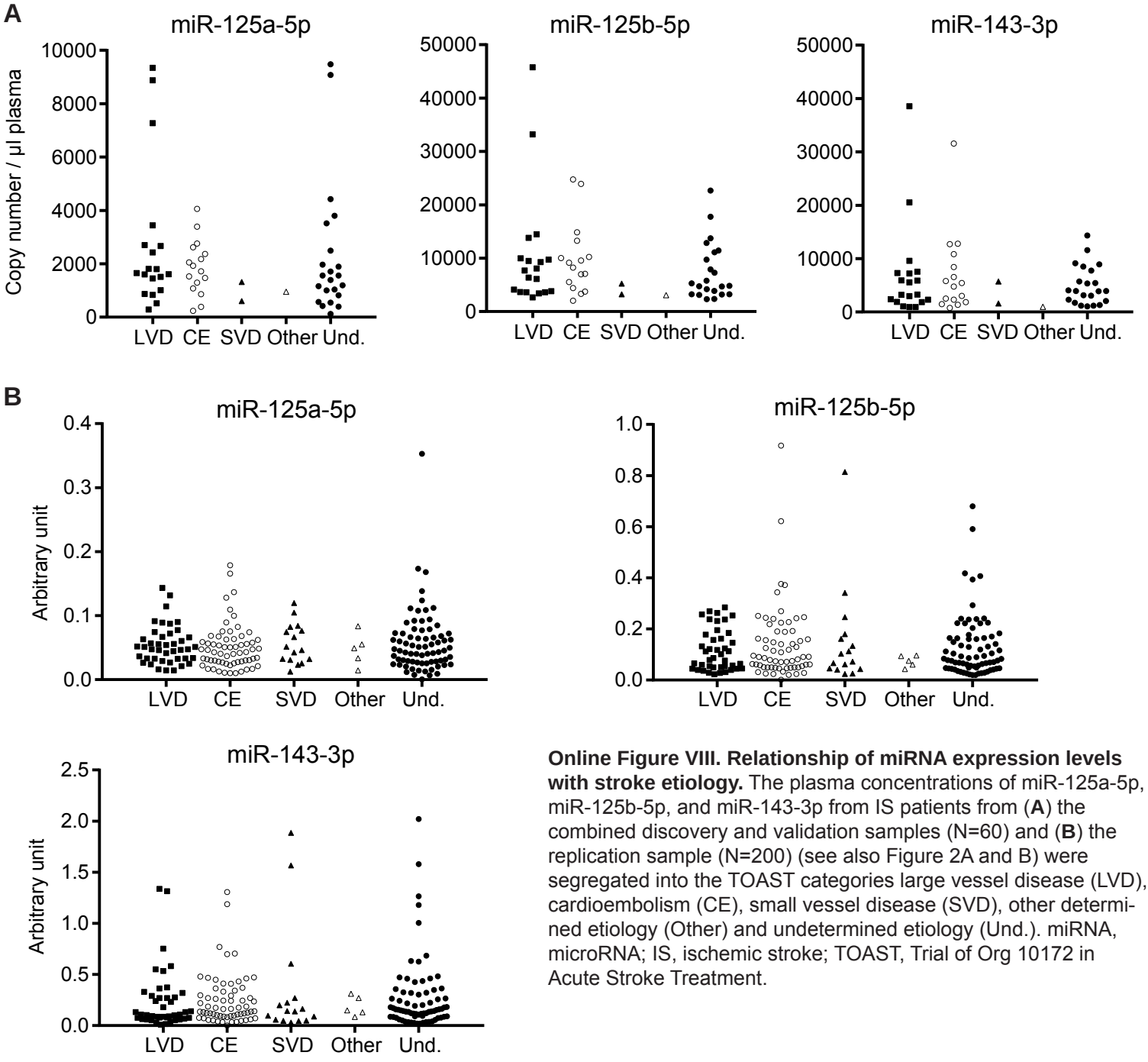


Online Figure VI. No expression bias by sequence similarity of miR-125a-5p and miR-125b-5p. (**A**) hsa-miR-125a-5p and hsa-miR-125b-5p have similar sequences. (**B**) Recombinant miR-125a-5p is not detected by the primers of miR-125b-5p and (**C**) vice versa. Mean ± SD. SD, standard deviation.

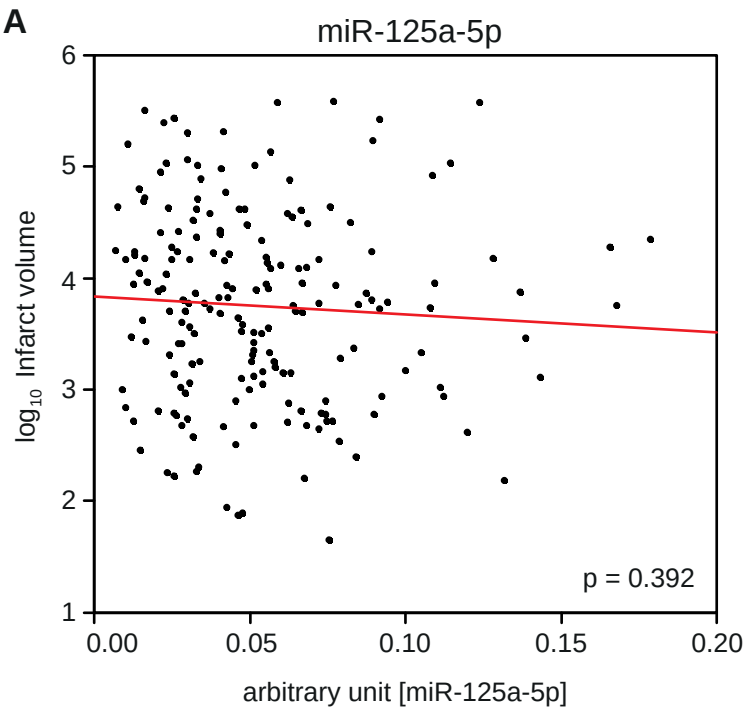


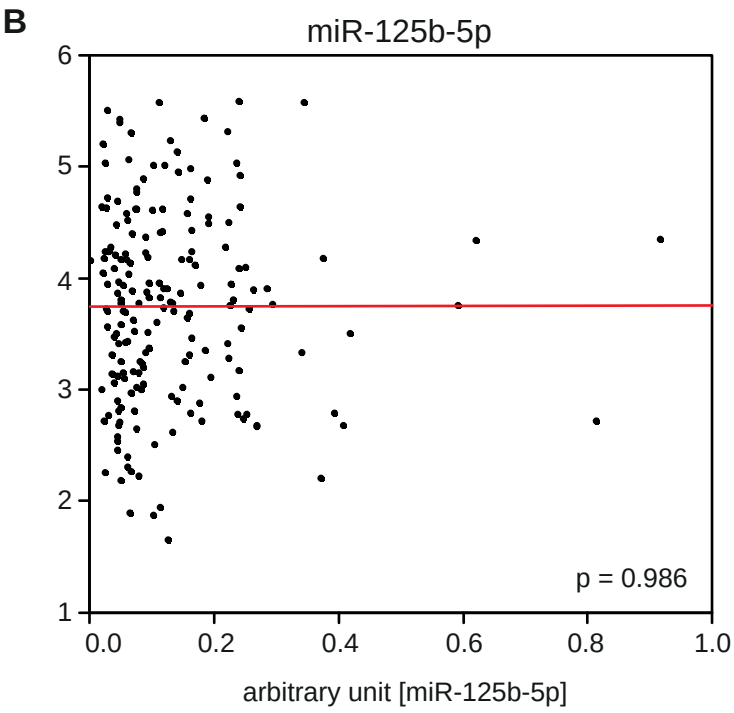


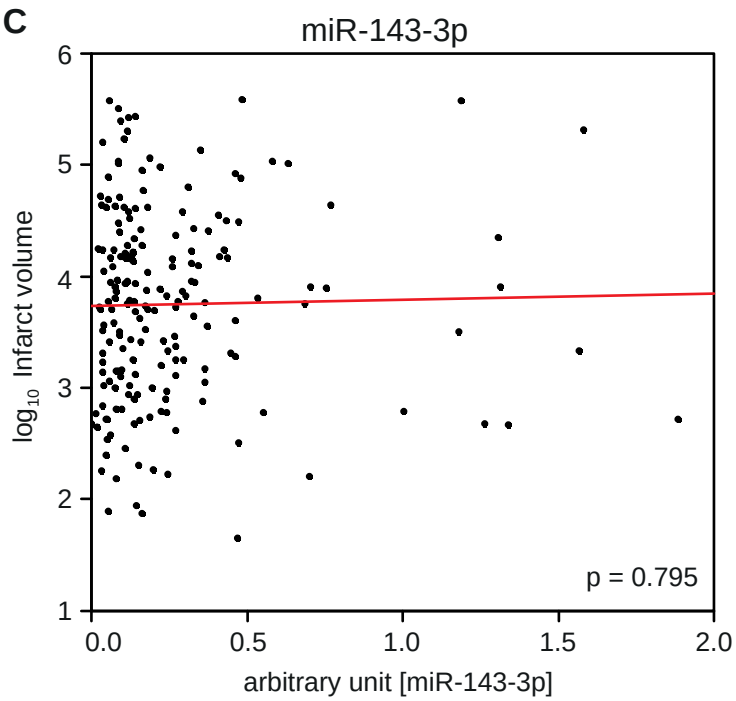
Online Figure VII. Relationship of miRNA expression levels with time since symptom onset. Fourth grade polynomial equations were used to plot expression levels of (**A**) miR-125a-5p, (**B**) miR-125b-5p, and (**C**) miR-143-3p from the 32 IS patients included in the longitudinal analysis at day one and two after symptom onset. For patients with unknown time from symptom onset, the difference between last seen healthy and first seen symptoms was calculated. The dashed line indicates the global maximum. Note that the second slight increase beyond 32 hours with few available measurements may result from statistical overfitting due to the use of fourth-grade polynomial equations. Polynomial regression analysis. miRNA, microRNA; IS, ischemic stroke; h, hours.



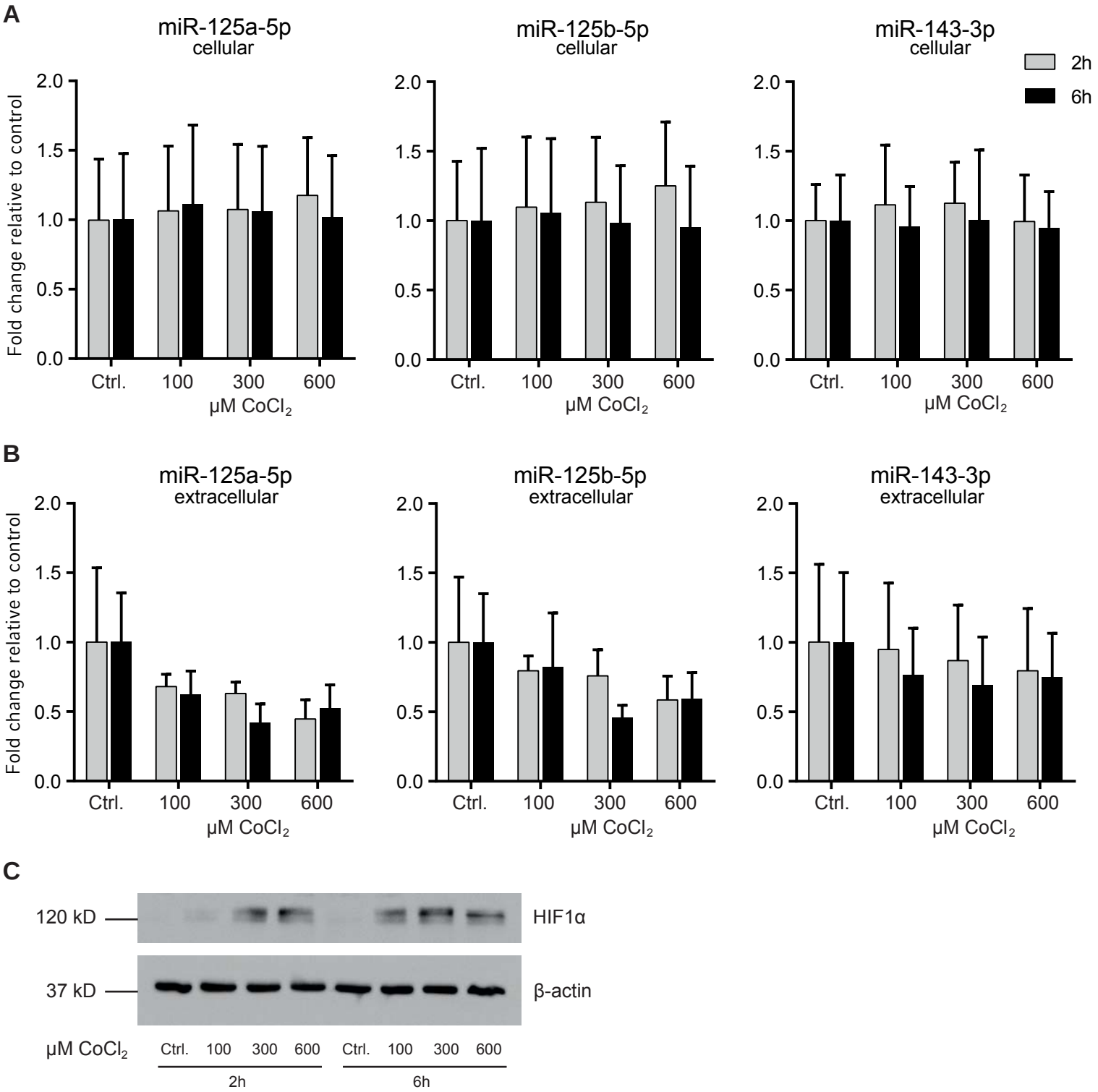
Online Figure VIII. Relationship of miRNA expression levels with stroke etiology. The plasma concentrations of miR-125a-5p, miR-125b-5p, and miR-143-3p from IS patients from (**A**) the combined discovery and validation samples (N=60) and (**B**) the replication sample (N=200) (see also Figure 2A and B) were segregated into the TOAST categories large vessel disease (LVD), cardioembolism (CE), small vessel disease (SVD), other determined etiology (Other) and undetermined etiology (Und.). miRNA, microRNA; IS, ischemic stroke; TOAST, Trial of Org 10172 in Acute Stroke Treatment.



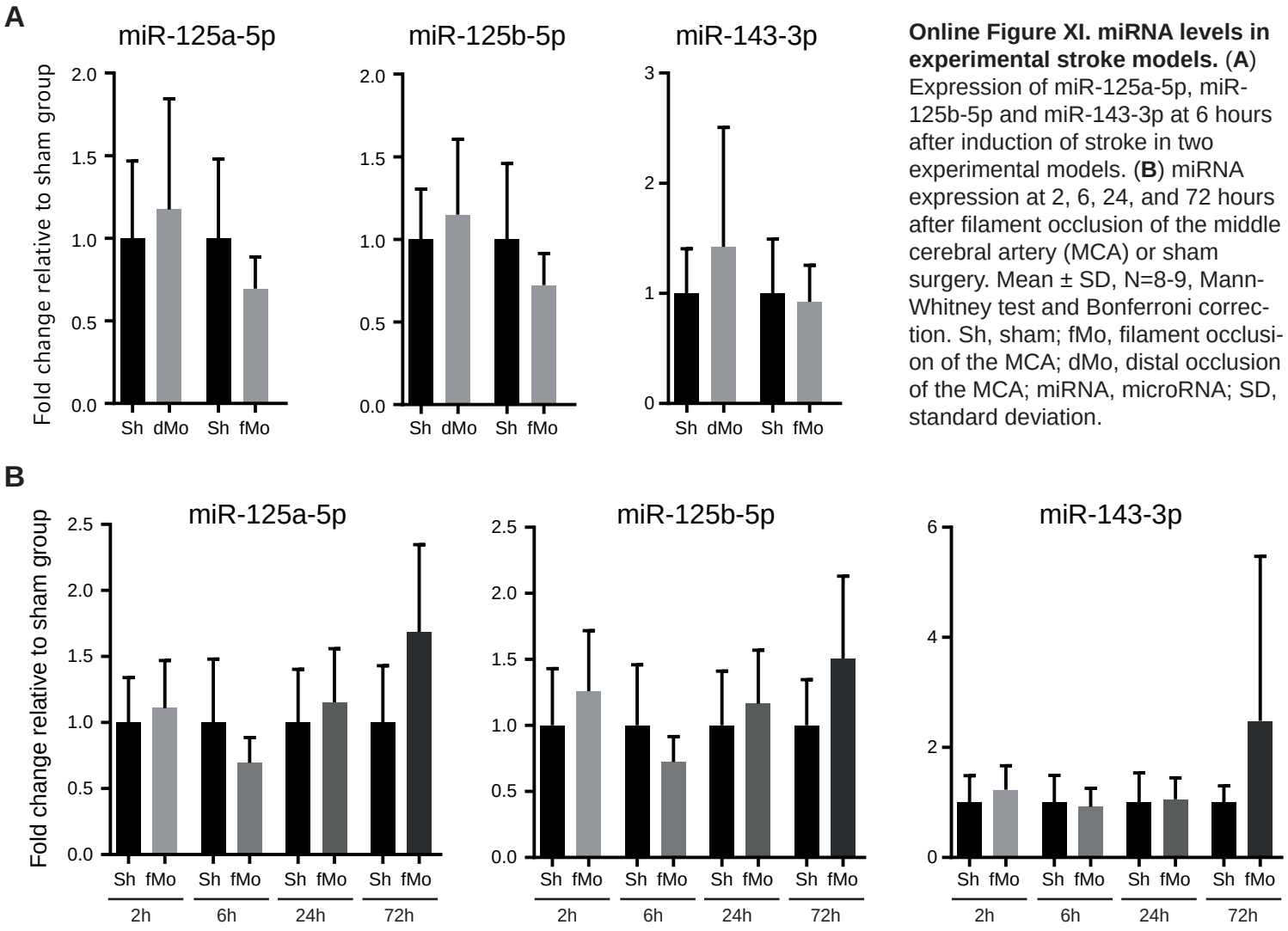




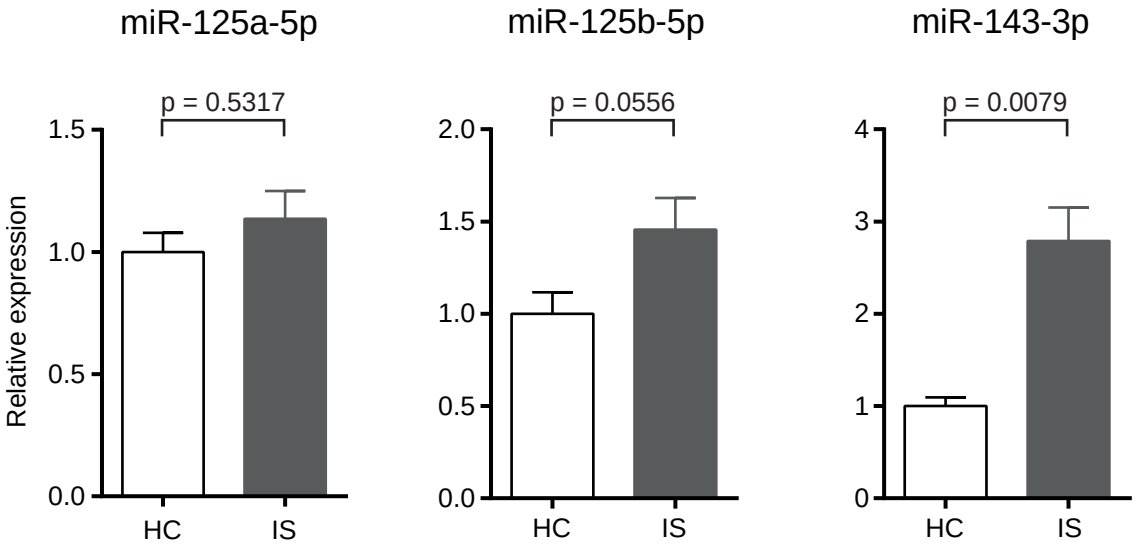
Online Figure IX. Relationship of miRNA expression levels with infarct volume. The transformed infarct volumes of ischemic stroke patients (N=188) were correlated with expression levels of (**A**) miR-125a-5p, (**B**) miR-125b-5p, and (**C**) miR-143-3p. Linear regression analysis. miRNA, microRNA.



Online Figure X. miRNA expression analyses after $CoCl_2$ challenge of N2a cells. (**A**) Cellular and (**B**) extracellular abundance of miR-125a-5p, miR-125b-5p and miR-143-3p did not change significantly after 2 and 6 hours of treatment with increasing concentrations (100 - 600 µM) of the HIF1α stabilizing agent $CoCl_2$ compared to control (H_2O) treated cells. (**C**) Expression of HIF1α is increased after $CoCl_2$ treatment. A and B: mean ± SD, Kruskal-Wallis-Test; C: western blot analysis. N2a, Neuro2a; $CoCl_2$, cobalt chloride; HIF1α, hypoxia-inducible factor 1-alpha; miRNA, microRNA; SD, standard deviation.



Online Figure XI. miRNA levels in experimental stroke models. (**A**) Expression of miR-125a-5p, miR-125b-5p and miR-143-3p at 6 hours after induction of stroke in two experimental models. (**B**) miRNA expression at 2, 6, 24, and 72 hours after filament occlusion of the middle cerebral artery (MCA) or sham surgery. Mean ± SD, N=8-9, Mann-Whitney test and Bonferroni correction. Sh, sham; fMo, filament occlusion of the MCA; dMo, distal occlusion of the MCA; miRNA, microRNA; SD, standard deviation.



Online Figure XII. Extracellular vesicle miRNA levels in IS patients compared to HCs. Expression levels of miR-125a-5p, miR-125b-5p, and miR-143-3p were analyzed in extracellular vesicles isolated from IS patients and HCs (in each group from 5 pools consisting each of 6 individual samples). Mean ± SEM, N=5 pools, Mann-Whitney test. Groups are indicated by: IS, ischemic stroke; and HC, healthy control subjects. miRNA, microRNA; SEM, standard error of the mean.

SUPPLEMENTAL TABLES

Online Table I: Individual sample characteristics

ID	group	total reads	miRNA reads	% miRNA	reads [miR-486 + miR-16-5p + miR-92a-3p] of miRNA reads (%)
1	stroke	6.644.136	889.364	13.4	43.5
2	stroke	7.151.832	409.431	5.7	47.7
3	stroke	6.500.093	142.132	2.2	42.3
4	stroke	8.318.539	2.796.547	33.6	50.1
5	stroke	8.960.549	2.155.168	24.1	47.1
6	stroke	8.019.565	247.076	3.1	41.9
7	stroke	6.967.596	1.075.369	15.4	47.0
8	stroke	15.279.030	446.267	2.9	44.0
9	stroke	6.720.124	428.881	6.4	45.9
10	stroke	7.672.286	582.012	7.6	46.7
11	stroke	5.999.103	122.117	2.0	40.7
12	stroke	8.086.114	152.341	1.9	43.9
13	stroke	8.512.960	223.141	2.6	39.2
14	stroke	8.518.381	204.232	2.4	43.2
15	stroke	6.839.795	137.378	2.0	48.3
16	stroke	4.282.503	228.649	5.3	45.5
17	stroke	9.878.416	1.458.512	14.8	48.9
18	stroke	8.638.248	193.398	2.2	43.5
19	stroke	9.051.766	189.795	2.1	41.3
20	stroke	8.161.355	122.554	1.5	44.6
21	control	9.530.966	1.165.867	12.2	49.4
22	control	9.450.249	816.919	8.6	48.9
23	control	8.894.061	4.513.161	50.7	49.6
24	control	11.508.674	4.603.933	40.0	49.6
25	control	9.601.776	1.730.942	18.0	51.4
26	control	9.161.124	4.517.736	49.3	52.0
27	control	8.553.193	880.956	10.3	47.5
28	control	8.797.517	378.127	4.3	45.5
29	control	7.680.805	770.032	10.0	47.2
30	control	7.471.164	923.621	12.4	47.8
31	control	7.731.336	48.057	0.6	43.7
32	control	6.419.856	392.607	6.1	44.6
33	control	6.771.409	98.856	1.5	47.2
34	control	2.430.061	88.400	3.6	48.5
35	control	3.220.558	307.606	9.6	51.4
36	control	2.061.139	51.510	2.5	50.0
37	control	377.400	17.409	4.6	43.7
38	control	1.452.484	25.602	1.8	44.8
39	control	11.030.262	1.025.496	9.3	45.7
40	control	10.834.234	1.336.697	12.3	47.3

The proportion of miRNA counts of total reads was 5.9 % (2.4 – 12.3 %; median and IQR). miRNA, microRNA; IQR, interquartile range.

Online Table III: Matching of initial evaluation of multimodal CT upon hospital arrival and delayed imaging

Patient_ID	Initial multimodal CT imaging upon hospital arrival			Delayed imaging resulting in final diagnosis	
	Non-contrast CT	CT angiography	CT perfusion	Modality	Final diagnosis
DS_01	Focal swelling in left frontoparietal lobe	Left MCA M2 segment occlusion	Mismatch in left MCA territory	MRI	Left MCA territory ischemic infarct
DS_02	Unremarkable	Unremarkable	Unremarkable	MRI	Left pons ischemic infarct
DS_03	Unremarkable	Unremarkable	Mismatch in right cerebellar hemisphere	MRI	Right cerebellar ischemic infarct
DS_04	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
DS_05	Unremarkable	Unremarkable	Unremarkable	MRI	Right MCA territory ischemic infarct
DS_06	Unremarkable	Unremarkable	Unremarkable	MRI	Right thalamus ischemic infarct
DS_07	Parenchymal hypoattenuation in right MCA territory	Right MCA M2 segment occlusion	Mismatch in right MCA territory	MRI	Right MCA territory ischemic infarct
DS_08	Unremarkable	Unremarkable	Mismatch in left MCA territory	MRI	Left MCA territory ischemic infarct
DS_09	Unremarkable	Unremarkable	Unremarkable	CT	Left MCA territory ischemic infarct
DS_10	Unremarkable	Unremarkable	Unremarkable	MRI	Left PCA territory ischemic infarct
DS_11	Unremarkable	Unremarkable	Unremarkable	MRI	Right MCA territory ischemic infarct
DS_12	Parenchymal hypoattenuation in left MCA territory	Unremarkable	Matched perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
DS_13	Unremarkable	Unremarkable	Unremarkable	MRI	Right ACA territory ischemic infarct

DS_14	Unremarkable	Left VA V3 occlusion	Unremarkable	MRI	Left/right VA/MCA territory ischemic infarcts
DS_15	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
DS_16	Parenchymal hypoattenuation in left MCA territory	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
DS_17	Unremarkable	Unremarkable	-	MRI	Left PCA territory ischemic infarct
DS_18	Hypoattenuation of the left caudate nucleus	Unremarkable	Perfusion defect between left MCA/ACA and left MCA/PCA	CT	Left MCA territory ischemic infarct
DS_19	Unremarkable	Unremarkable	Mismatch in left MCA territory	MRI	Left MCA and right PCA ischemic infarcts
DS_20	Unremarkable	Right MCA M2 segment occlusion	Mismatch in right MCA territory	MRI	Right MCA territory ischemic infarct
VS_01	Unremarkable	Unremarkable	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
VS_02	Parenchymal hypoattenuation in right MCA territory	Right ICA and MCA occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
VS_03	Unremarkable	Unremarkable	Matched perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
VS_04	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
VS_05	Unremarkable	Possible left MCA M4 segment occlusion	Mismatch in left MCA territory	MRI	Left/right MCA territory ischemic infarcts
VS_06	Unremarkable	Unremarkable	Unremarkable	MRI	Left pons ischemic infarct
VS_07	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
VS_08	Unremarkable	Left MCA M2 segment occlusion	Mismatch in left MCA territory	MRI	Left MCA territory ischemic infarct

VS_09	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
VS_10	Unremarkable	Unremarkable	Unremarkable	MRI	Right ACA territory ischemic infarct
VS_11	Unremarkable	Unremarkable	Unremarkable	CT	Left MCA territory ischemic infarct
VS_12	Unremarkable	Left VA V2 segment occlusion	Unremarkable	MRI	Right SCA territory ischemic infarct
VS_13	Parenchymal hypoattenuation in right ACA/MCA territories	Unremarkable	Unremarkable	MRI	Right MCA territory ischemic infarct
VS_14	Unremarkable	Unremarkable	Unremarkable	CT	Left MCA territory ischemic infarct
VS_15	Parenchymal hypoattenuation in left MCA territory	Unremarkable	Unremarkable	CT	Left MCA territory ischemic infarct
VS_16	Unremarkable	Unremarkable	Unremarkable	MRI	Right pons ischemic infarct
VS_17	Unremarkable	Unremarkable	Unremarkable	CT	Left MCA territory ischemic infarct
VS_18	Unremarkable	Unremarkable	-	CT	Left MCA territory ischemic infarct
VS_19	Unremarkable	Unremarkable	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
VS_20	Parenchymal hypoattenuation in left PCA territory	Unremarkable	Matched perfusion defect within the left PCA territory	MRI	Left PCA territory ischemic infarct
VS_21	Unremarkable	Unremarkable	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
VS_22	Unremarkable	Left MCA M1/2 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
VS_23	Unremarkable	Right MCA M2 segment occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct

VS_24	Unremarkable	Unremarkable	Unremarkable	MRI	Left thalamus ischemic infarct
VS_25	Hyperdense MCA sign (left)	Left MCA M2 segment stenosis	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
VS_26	Unremarkable	Right PCA P1/2 segment occlusion	Mismatch in right PCA territory	MRI	Right PCA territory ischemic infarct
VS_27	Unremarkable	Left MCA M1 segment stenosis	Mismatch in left MCA territory	CT	Left MCA territory ischemic infarct
VS_28	Parenchymal hypoattenuation in right MCA territory	Right MCA M1 segment stenosis	-	MRI	Right MCA territory ischemic infarct
VS_29	Unremarkable	Unremarkable	Unremarkable	MRI	Right thalamus ischemic infarct
VS_30	Unremarkable	Unremarkable	Unremarkable	MRI	Right PICA territory ischemic infarct
VS_31	Parenchymal hypoattenuation of right inferior frontal gyrus	Possible right MCA M3 segment occlusion	Matched perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
VS_32	Unremarkable	Left ICA stenosis	Unremarkable	MRI	Left MCA territory ischemic infarct
VS_33	Unremarkable	Left ICA stenosis	Unremarkable	MRI	Left MCA territory ischemic infarct
VS_34	Unremarkable	-	Matched perfusion defect in right PICA territory	CT	Right PICA territory ischemic infarct
VS_35	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
VS_36	Unremarkable	Unremarkable	-	MRI	Right MCA territory ischemic infarct
VS_37	Unremarkable	Unremarkable	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
VS_38	Unremarkable	Left MCA M1 segment occlusion	Mismatch in left MCA territory	MRI	Left MCA territory ischemic infarct

VS_39	Unremarkable	Left ICA occlusion	Perfusion defect in left PCA territory	CT	Left MCA/PCA territory ischemic infarcts
VS_40	Parenchymal hypoattenuation in right parietal lobe	Unremarkable	Matched perfusion defect in right parieto-occipital region	MRI	Right MCA territory ischemic infarct
RS_1	Unremarkable	Right VA V3 segment occlusion	-	MRI	Right SCA territory ischemic infarct
RS_2	Unremarkable	Unremarkable	-	MRI	Left pontine ischemic infarct
RS_3	Hyperdense MCA sign (right)	-	-	CT	Right MCA territory ischemic infarct
RS_4	Parenchymal hypoattenuation of the left basal ganglia	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_5	Parenchymal hypoattenuation of the left insula and basal ganglia	Left ICA occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_6	Parenchymal hypoattenuation in the left frontal lobe	Unremarkable	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_7	Parenchymal hypoattenuation in the right temporal lobe	Right ICA occlusion	Unremarkable	MRI	Right MCA territory ischemic infarct
RS_8	Parenchymal hypoattenuation of the left insula and basal ganglia	Left MCA M3 segment occlusion	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_9	Focal swelling in left parietal lobe	Left ICA thrombus	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_10	Unremarkable	Unremarkable	Unremarkable	MRI	Left and right MCA territory ischemic infarcts
RS_11	Unremarkable	Unremarkable	Perfusion defect in right ACA territory	CT	Right ACA territory ischemic infarct
RS_12	Unremarkable	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_13	Parenchymal hypoattenuation of the left internal capsule	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct

RS_14	Unremarkable	Left ICA occlusion	Unremarkable	MRI	Left ACA and MCA territory ischemic infarcts
RS_15	Parenchymal hypoattenuation of the left insula and basal ganglia	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_16	Unremarkable	Unremarkable	-	MRI	Right MCA territory ischemic infarct
RS_17	Unremarkable	Left MCA M3 segment occlusion	Perfusion defect in left MCA territory	MRI	Left MCA and PICA territory ischemic infarcts
RS_18	Unremarkable	Right PCA P2 segment occlusion	Perfusion defect in right PCA territory	CT	Right PCA territory ischemic infarct
RS_19	Unremarkable	Right MCA M1 segment occlusion	Perfusion defect in right MCA territory	MRI	Right MCA, left MCA/PCA territory ischemic infarcts
RS_20	Unremarkable	Bilateral PCA stenoses	-	MRI	Right PCA territory ischemic infarct
RS_21	Unremarkable	Left MCA M3 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_22	Unremarkable	Right (older) ICA occlusion	Perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_23	Unremarkable	Left PCA stenosis	-	MRI	Left thalamic ischemic infarct
RS_24	Unremarkable	Right ICA occlusion	-	MRI	Right ACA territory ischemic infarct
RS_25	Unremarkable	Left ICA occlusion	Perfusion defect in right MCA territory	MRI	Right/left MCA and ACA territory ischemic infarcts
RS_26	Unremarkable	Right MCA M2 segment occlusion	Perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_27	Hyperdense MCA sign (left)	Left MCA M1 segment occlusion	-	CT	Left MCA territory ischemic infarct
RS_28	Unremarkable	Left MCA M2 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct

RS_29	Parenchymal hypoattenuation of the left insula	-	-	CT	Left MCA territory ischemic infarct
RS_30	Unremarkable	Right ICA stenosis	-	MRI	Right MCA territory ischemic infarct
RS_31	Unremarkable	Right ICA and MCA occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_32	Parenchymal hypoattenuation of the left parietal lobe	Left ICA stenosis	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_33	Parenchymal hypoattenuation of the left insula	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_34	Parenchymal hypoattenuation of the left basal ganglia	Left MCA M1 stenosis	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_35	Unremarkable	-	-	MRI	Right MCA territory ischemic infarct
RS_36	Unremarkable	Unremarkable	Perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_37	Unremarkable	Unremarkable	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_38	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
RS_39	Unremarkable	Unremarkable	Unremarkable	CT	Left MCA territory ischemic infarct
RS_40	Unremarkable	Left MCA M3 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_41	Parenchymal hypoattenuation in the right SCA territory	Right MCA M1 segment occlusion	Perfusion defect in right MCA territory	CT	Right MCA and SCA territory ischemic infarcts
RS_42	Unremarkable	Unremarkable	Unremarkable	MRI	Right pontine ischemic infarct
RS_43	Parenchymal hypoattenuation of the right insula	Right MCA M2 segment occlusion	-	CT	Right MCA territory ischemic infarct

RS_44	Parenchymal hypoattenuation of the left MCA territory	Left ICA and MCA M1 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_45	Unremarkable	-	-	CT	Right MCA territory ischemic infarct
RS_46	Parenchymal hypoattenuation of the right basal ganglia	-	-	CT	Right MCA territory ischemic infarct
RS_47	Unremarkable	Unremarkable	Unremarkable	MRI	Right MCA territory ischemic infarct
RS_48	Parenchymal hypoattenuation in the right temporal lobe	Right MCA M1 segment occlusion	Perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_49	Parenchymal hypoattenuation of the left insula	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_50	Parenchymal hypoattenuation in the right MCA territory	Right ICA occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_51	Parenchymal hypoattenuation of the left MCA territory	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_52	Unremarkable	Unremarkable	Unremarkable	MRI	Left thalamic/capsule ischemic infarct
RS_53	Parenchymal hypoattenuation of the left frontal lobe	Unremarkable	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_54	Parenchymal hypoattenuation of the right frontal lobe	Right MCA M1 segment occlusion	Perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_55	Parenchymal hypoattenuation of the right temporal lobe	Unremarkable	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_56	Unremarkable	Unremarkable	Unremarkable	MRI	Left pontine ischemic infarct
RS_57	Parenchymal hypoattenuation of the right temporal lobe	Right MCA M2 segment occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_58	Parenchymal hypoattenuation of the left frontal lobe	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct

RS_59	Parenchymal hypoattenuation in the right MCA territory	Right ICA occlusion	Perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_60	Parenchymal hypoattenuation of the left insula	Left ICA occlusion	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_61	Parenchymal hypoattenuation in the right MCA territory	Right MCA M1 segment occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_62	Parenchymal hypoattenuation of the right basal ganglia	Unremarkable	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_63	Unremarkable	Unremarkable	-	MRI	Right MCA territory ischemic infarct
RS_64	Parenchymal hypoattenuation in the right PCA territory	Right PCA P4 segment occlusion	-	MRI	Right PCA territory ischemic infarct
RS_65	Unremarkable	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	MRI	Left MCA and PICA territory ischemic infarcts
RS_66	Parenchymal hypoattenuation in the right MCA territory	Unremarkable	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_67	Parenchymal hypoattenuation of the right MCA territory	Right ICA occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_68	Parenchymal hypoattenuation in the left MCA territory	Left ICA stenosis	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_69	Unremarkable	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_70	Hyperdense MCA sign (left)	Left MCA M2 segment occlusion	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_71	Parenchymal hypoattenuation in the left MCA territory	Left MCA M1 segment occlusion	-	CT	Left MCA territory ischemic infarct
RS_72	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
RS_73	Parenchymal hypoattenuation in the right MCA territory	Right MCA M2 segment occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct

RS_74	Parenchymal hypoattenuation in the left basal ganglia	Unremarkable	Unremarkable	CT	Left MCA territory ischemic infarct
RS_75	Parenchymal hypoattenuation in the right MCA territory	Right ICA occlusion	Perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_76	Unremarkable	Left MCA M2 segment stenosis	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_77	Parenchymal hypoattenuation in the right MCA territory	Right MCA M1 segment occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_78	Unremarkable	Unremarkable	-	MRI	Left MCA territory ischemic infarct
RS_79	Unremarkable	Unremarkable	Unremarkable	MRI	Bilateral thalamic ischemic infarcts
RS_80	Parenchymal hypoattenuation in the left MCA territory	Left MCA M3 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_81	Unremarkable	Unremarkable	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_82	Unremarkable	-	-	MRI	Left PCA territory ischemic infarct
RS_83	Hyperdense MCA sign (left)	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_84	Parenchymal hypoattenuation in the left MCA territory	Left MCA M3 segment occlusion	-	CT	Left MCA territory ischemic infarct
RS_85	Parenchymal hypoattenuation in the right MCA territory	Right MCA M2 segment occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_86	Unremarkable	Left PCA P1 segment occlusion	Perfusion defect in left PCA territory	CT	Left PCA and SCA territory ischemic infarcts
RS_87	Parenchymal hypoattenuation in the left temporal lobe	Left ICA and MCA occlusion	Perfusion defect in left MCA territory	MRI	Left MCA and PCA territory ischemic infarcts
RS_88	Parenchymal hypoattenuation in the left MCA territory	-	-	CT	Left MCA territory ischemic infarct

RS_89	Parenchymal hypoattenuation in the left frontal lobe	Left CCA, ICA and MCA occlusion	Perfusion defect in left ACA and MCA territories	CT	Left MCA and ACA territory ischemic infarcts
RS_90	- (external MRI acquisition)	-	-	MRI	Left MCA territory ischemic infarct
RS_91	Parenchymal hypoattenuation in the left basal ganglia	Left MCA M1 segment occlusion	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_92	Parenchymal hypoattenuation in the left MCA territory	Left ICA occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_93	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
RS_94	Parenchymal hypoattenuation in the right PCA territory	Unremarkable	-	MRI	Right PCA territory ischemic infarct
RS_95	Parenchymal hypoattenuation in the right frontal lobe	Right MCA M2 segment occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_96	Unremarkable	Unremarkable	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_97	Unremarkable	Left MCA M3 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_98	Parenchymal hypoattenuation in the left cerebellar hemipshere	Right and left VA V4 segment occlusion	Perfusion defect in left AICA and PICA territories	MRI	Left PICA territory ischemic infarct
RS_99	Unremarkable	Left PCA P2 segment occlusion	Perfusion defect in left PCA territory	MRI	Left PCA territory ischemic infarct
RS_100	Parenchymal hypoattenuation in the left pulvinar	Left PCA P1 segment occlusion	Perfusion defect in left PCA territory	MRI	Left PCA territory ischemic infarct
RS_101	Unremarkable	Unremarkable	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_102	Unremarkable	Right MCA M1 segment occlusion	Perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_103	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct

RS_104	Unremarkable	Basilar artery thrombosis	Mismatch in left cerebellar hemisphere	MRI	Bilateral PCA and left SCA territory ischemic infarcts
RS_105	Unremarkable	Unremarkable	-	CT	Left pontine ischemic infarct
RS_106	Unremarkable	Left MCA M1 segment occlusion	-	MRI	Left MCA territory ischemic infarct
RS_107	Unremarkable	Unremarkable	-	MRI	Right MCA territory ischemic infarct
RS_108	Unremarkable	Left VA V4 segment stenosis	Unremarkable	MRI	Left medulla oblongata ischemic infarct
RS_109	Unremarkable	Unremarkable	-	CT	Left MCA territory ischemic infarct
RS_110	Unremarkable	Unremarkable	-	MRI	Left PCA territory ischemic infarct
RS_111	Unremarkable	Right ICA occlusion	Mismatch in right ACA territory	CT	Right ACA territory ischemic infarct
RS_112	Unremarkable	-	-	MRI	Right MCA territory ischemic infarct
RS_113	Parenchymal hypoattenuation in the right MCA territory	Right ICA occlusion	Matched perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_114	Unremarkable	Unremarkable	-	CT	Right MCA territory ischemic infarct
RS_115	Unremarkable	Unremarkable	-	MRI	Left MCA and PCA territory ischemic infarcts
RS_116	Unremarkable	Right MCA M2 segment occlusion	Mismatch in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_117	Parenchymal hypoattenuation in the left basal ganglia	Left MCA M2 segment occlusion	Mismatch in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_118	Unremarkable	Left ICA occlusion/ high-grade stenosis	Mismatch in left MCA territory	CT	Left MCA territory ischemic infarct

RS_119	Unremarkable	-	-	MRI	Left MCA territory ischemic infarct
RS_120	Unremarkable	Right VA V4 segment occlusion	Matched perfusion defect in right PICA territory	MRI	Right PICA territory ischemic infarct
RS_121	Parenchymal hypoattenuation in the right MCA territory	Right MCA M2 segment occlusion	-	CT	Right MCA territory ischemic infarct
RS_122	Unremarkable	Unremarkable	Unremarkable	MRI	Right thalamic ischemic infarct
RS_123	Unremarkable	Right MCA M1 segment occlusion	Mismatch in right MCA territory	CT	Right MCA territory ischemic infarct
RS_124	Unremarkable	Unremarkable	-	MRI	Bilateral PCA territory ischemic infarcts
RS_125	Infarction of left basal ganglia with hemorrhagic transformation	Unremarkable	-	MRI	Bilateral MCA territory ischemic infarcts
RS_126	Unremarkable	Unremarkable	Unremarkable	CT	Left MCA territory ischemic infarct
RS_127	Parenchymal hypoattenuation of the left insula	Left ICA and MCA M1 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_128	Unremarkable	Left MCA M2 segment occlusion	Mismatch in left MCA territory	CT	Left MCA territory ischemic infarct
RS_129	Unremarkable	-	-	MRI	Left MCA territory ischemic infarct
RS_130	Unremarkable	Unremarkable	Unremarkable	MRI	Right MCA territory ischemic infarct
RS_131	Unremarkable	Right MCA M2 segment occlusion	Mismatch in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_132	Unremarkable	Unremarkable	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_133	Unremarkable	Left MCA M3 segment occlusion	Perfusion defect in left MCA territory	MRI	Left MCA territory ischemic infarct

RS_134	Unremarkable	Unremarkable	Unremarkable	MRI	Left thalamic ischemic infarct
RS_135	Unremarkable	Left MCA M2 segment occlusion	Mismatch in left MCA territory	CT	Left MCA territory ischemic infarct
RS_136	Unremarkable	Unremarkable	-	CT	Left MCA territory ischemic infarct
RS_137	Parenchymal hypoattenuation of the right frontal lobe	Unremarkable	Perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_138	Unremarkable	Unremarkable	Unremarkable	MRI	Right MCA territory ischemic infarct
RS_139	Unremarkable	Unremarkable	Unremarkable	MRI	Right medulla oblongata ischemic infarct
RS_140	Parenchymal hypoattenuation in the left MCA territory	Unremarkable	-	MRI	Left MCA territory ischemic infarct
RS_141	Parenchymal hypoattenuation in the right frontal lobe	Right MCA M3 segment occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_142	Unremarkable	Right MCA M2 segment occlusion	Mismatch in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_143	Parenchymal hypoattenuation in the left PCA territory	Unremarkable	-	CT	Left PCA and MCA territory ischemic infarcts
RS_144	Unremarkable	Left MCA M2 segment occlusion	Perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_145	Parenchymal hypoattenuation in the left insula	Left ICA and MCA M1 segment occlusion	Matched perfusion defect in left ACA and MCA territory	CT	Left ACA and MCA territory ischemic infarcts
RS_146	Parenchymal hypoattenuation in the left MCA territory	Left MCA M1 segment occlusion	Mismatch in left MCA territory	CT	Left MCA territory ischemic infarct
RS_147	Unremarkable	Right MCA M1 segment occlusion	Mismatch in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_148	Parenchymal hypoattenuation in the right occipital lobe	Right PCA P3 segment occlusion	Perfusion defect in right PCA territory	MRI	Right PCA territory ischemic infarct

RS_149	Parenchymal hypoattenuation of the right MCA territory	Right MCA M1 segment occlusion	Perfusion defect in right MCA territory	CT	Right MCA territory ischemic infarct
RS_150	Parenchymal hypoattenuation in the left MCA territory	Left MCA M1 segment occlusion	Perfusion defect in right MCA territory	MRI	Left MCA territory ischemic infarct
RS_151	Parenchymal hypoattenuation of the left insula	Left MCA M1 segment occlusion	Matched perfusion defect in left MCA territory	CT	Left MCA territory ischemic infarct
RS_152	Unremarkable	Unremarkable	-	MRI	Left MCA territory ischemic infarct
RS_153	Parenchymal hypoattenuation of the left occipital lobe	Unremarkable	Matched perfusion defect in left PCA territory	CT	Left PCA territory ischemic infarct
RS_154	Parenchymal hypoattenuation of the right basal ganglia	Right MCA M1 segment occlusion	Mismatch in right MCA territory	CT	Right MCA territory ischemic infarct
RS_155	Unremarkable	Unremarkable	-	MRI	Right thalamic ischemic infarct
RS_156	Parenchymal hypoattenuation of the left temporal lobe	Left MCA M1 segment occlusion	-	MRI	Left MCA territory ischemic infarct
RS_157	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
RS_158	Unremarkable	Unremarkable	-	CT	Left MCA territory ischemic infarct
RS_159	Unremarkable	Unremarkable	-	MRI	Left pontine ischemic infarct
RS_160	Unremarkable	Left MCA M3 segment occlusion	Mismatch in left MCA territory	CT	Left MCA territory ischemic infarct
RS_161	Parenchymal hypoattenuation of the left PCA territory	Left PCA P1 segment occlusion	-	CT	Left PCA territory ischemic infarct
RS_162	Unremarkable	Unremarkable	Mismatch in right MCA territory	MRI	Right PCA and MCA territory ischemic infarcts
RS_163	Parenchymal hypoattenuation in the right frontal lobe	Right ICA occlusion	-	MRI	Right MCA territory ischemic infarct

RS_164	Parenchymal hypoattenuation in the left temporal lobe	Unremarkable	-	CT	Left MCA territory ischemic infarct
RS_165	Unremarkable	Basilar artery thrombosis	Perfusion defect in right PCA and bilateral SCA territories	CT	Right PCA and bilateral SCA territory ischemic inf.
RS_166	Parenchymal hypoattenuation in the left temporal lobe	Unremarkable	-	MRI	Left MCA territory ischemic infarct
RS_167	Parenchymal hypoattenuation in the right basal ganglia	Right MCA M1 segment occlusion	Mismatch in right MCA territory	CT	Right MCA territory ischemic infarct
RS_168	Unremarkable	Right MCA M2 segment occlusion	Mismatch in right MCA territory	CT	Right MCA territory ischemic infarct
RS_169	Parenchymal hypoattenuation in the left frontal lobe	Unremarkable	Mismatch in right MCA territory	MRI	Left MCA territory ischemic infarct
RS_170	Unremarkable	Unremarkable	Mismatch in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_171	Unremarkable	Right PCA P3 segment occlusion	Perfusion defect in right parietal lobe	MRI	Right MCA and PCA territory ischemic infarcts
RS_172	Unremarkable	Unremarkable	-	MRI	Right PICA territory ischemic infarct
RS_173	Unremarkable	Right ICA stenosis	-	MRI	Right MCA territory ischemic infarct
RS_174	Unremarkable	Unremarkable	-	MRI	Left MCA territory ischemic infarct
RS_175	Parenchymal hypoattenuation in the right frontal lobe	Unremarkable	-	CT	Right MCA territory ischemic infarct
RS_176	Parenchymal hypoattenuation in the right MCA territory	Unremarkable	-	MRI	Right MCA territory ischemic infarct
RS_177	Unremarkable	Left MCA M2 segment occlusion	Mismatch in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_178	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct

RS_179	Unremarkable	Unremarkable	Mismatch in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_180	Parenchymal hypoattenuation of the right basal ganglia	Right MCA M1 segment occlusion	-	CT	Right MCA territory ischemic infarct
RS_181	Unremarkable	Left PCA stenosis	-	MRI	Bilateral mesencephalic ischemic infarcts
RS_182	Unremarkable	Unremarkable	-	MRI	Right MCA territory ischemic infarct
RS_183	Unremarkable	Left MCA M2 segment occlusion	Mismatch in left ACA and MCA territories	MRI	Left ACA and MCA territory ischemic infarcts
RS_184	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct
RS_185	Parenchymal hypoattenuation in the left frontal lobe	Unremarkable	Perfusion defect in left ACA territory	CT	Left ACA territory ischemic infarct
RS_186	Parenchymal hypoattenuation in the left MCA territory	Left CCA occlusion	Matched perfusion defect in left MCA territory	CT	Left ACA and MCA territory ischemic infarcts
RS_187	Unremarkable	Unremarkable	Unremarkable	MRI	Left pontine ischemic infarct
RS_188	Parenchymal hypoattenuation in the left basal ganglia	Left ICA occlusion	-	MRI	Left MCA territory ischemic infarct
RS_189	Parenchymal hypoattenuation in the right ACA territory	Right ACA A2 segment occlusion	Mismatch in right ACA territory	CT	Right ACA territory ischemic infarct
RS_190	Parenchymal hypoattenuation in the left MCA territory	Left ICA occlusion	Mismatch in left MCA territory	MRI	Left MCA territory ischemic infarct
RS_191	Parenchymal hypoattenuation in the left temporal lobe	Left ICA stenosis	-	MRI	Bilateral MCA territory ischemic infarcts
RS_192	Unremarkable	Basilar artery thrombosis	-	MRI	Right PCA territory ischemic infarct
RS_193	Unremarkable	Unremarkable	Unremarkable	MRI	Left MCA territory ischemic infarct

RS_194	Unremarkable	Unremarkable	Matched perfusion defect in right MCA territory	MRI	Right MCA territory ischemic infarct
RS_195	- (immediate MRI acquisition)	-	-	MRI	Right MCA territory ischemic infarct
RS_196	Unremarkable	Unremarkable	-	MRI	Left medulla oblongata ischemic infarct
RS_197	- (external MRI acquisition)	-	-	MRI	Right MCA territory ischemic infarct
RS_198	Unremarkable	Left MCA M1 segment occlusion	-	MRI	Bilateral MCA territory ischemic infarcts
RS_199	Parenchymal hypoattenuation in the left MCA territory	Left MCA M1 segment occlusion	-	MRI	Left MCA territory ischemic infarct
RS_200	Parenchymal hypoattenuation in the left MCA territory	Unremarkable	-	MRI	Left MCA territory ischemic infarct

CT, computed tomography; MRI, magnetic resonance imaging; ICA, internal carotid artery; MCA, middle cerebral artery; ACA, anterior cerebral artery; PCA, posterior cerebral artery; VA, vertebral artery; PICA, posterior inferior cerebellar artery; SCA, superior cerebellar artery; unremarkable, no signs for infarction; mismatch/match, with reference to blood flow and blood volume. Light grey indicates signs of infarction in at least one CT modality matching the final diagnosis.

SUPPLEMENTAL REFERENCES

1. Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE, 3rd. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke; a journal of cerebral circulation.* 1993;24:35-41.

2. Kaudewitz D, Skroblin P, Bender LH, et al. Association of MicroRNAs and YRNAs With Platelet Function. *Circulation research.* 2016;118:420-432.

3. McAlexander MA, Phillips MJ, Witwer KW. Comparison of Methods for miRNA Extraction from Plasma and Quantitative Recovery of RNA from Cerebrospinal Fluid. *Frontiers in genetics.* 2013;4:83.

4. Kaudewitz D, Lee R, Willeit P, McGregor R, Markus HS, Kiechl S, Zampetaki A, Storey RF, Channon KM, Mayr M. Impact of intravenous heparin on quantification of circulating microRNAs in patients with coronary artery disease. *Thrombosis and haemostasis.* 2013;110:609-615.

5. Boeckel JN, Thome CE, Leistner D, Zeiher AM, Fichtlscherer S, Dimmeler S. Heparin selectively affects the quantification of microRNAs in human blood samples. *Clinical chemistry.* 2013;59:1125-1127.

6. Liaw A, Wiener M. Classification and Regression by randomForest. *R News.* 2002;2:18-22.

7. Venables W, Ripley B. Modern Applied Statistics with S. 2002.

8. Team RC. R: A Language and Environment for Statistical Computing. 2017.

4.2. Agreement between TOAST and CCS ischemic stroke classification: the NINDS SiGN Study

Correct determination of stroke etiology is key to guide secondary prevention of IS patients. Lacking a tissue-based biomarker or a single diagnostic test with high accuracy, current classifications systems are limited to integrating findings from various technical studies. Here, we aimed at determining the overlap of two well-established systems, the Trial of Org 10172 in Acute Stroke Treatment (TOAST) and the more recently developed Causative Classification of Stroke (CCS), for classifying IS.

16.267 IS patients from 10 European and 10 US sites were enrolled. With varying agreement between study sites, the overall agreement between CCS and TOAST was moderate (κ = 0.59, 95 % CI, 0.58 – 0.60). The agreement on specific subtypes was higher for large-artery atherosclerosis (κ = 0.71) compared to small-artery occlusions (κ = 0.56). Although both systems idenfied a nearly identical number of patients as undeterminable (CCS: 4.673 cases; TOAST: 4.664 cases), the agreement was only κ = 0.44 (95 % CI 0.43 – 0.46).

In conclusion, the agreement between TOAST and CCS is lower than previously reported. Both classification systems leave more than 25 % of IS patients undetermined. Further studies are warranted to extend the biomarker portfolio for the determination of stroke etiology beyond clinical and imaging findings.

Author contribution: recruitment of IS patients, classification of IS patients according to their etiology (please see section 10 for further details).

Agreement between TOAST and CCS ischemic stroke classification

The NINDS SiGN Study

Patrick F. McArdle, PhD
Steven J. Kittner, MD
Hakan Ay, MD
Robert D. Brown, Jr., MD
James F. Meschia, MD
Tatjana Rundek, MD
Sylvia Wassertheil-Smoller, PhD
Daniel Woo, MD
Gunnar Andsberg, MD
Alessandro Biffi, MD
David A. Brenner, MD, MS
John W. Cole, MD, MS
Roderick Corriveau, PhD
Paul I.W. de Bakker, PhD
Hossein Delavaran, MD
Martin Dichgans, MD
Raji P. Grewal, MD
Katrina Gwinn, MD
Mohammed Huq, MD
Christina Jern, MD
Jordi Jimenez-Conde, MD, PhD
Katarina Jood, MD
Robert C. Kaplan, PhD
Petra Katschnig, MD
Michael Katsnelson, MD
Daniel L. Labovitz, MD
Robin Lemmens, MD
Linxin Li, MD
Arne Lindgren, MD
Hugh S. Markus, FRCP
Leema R. Peddareddygari, MD
Annie Pedersén, MD
Joanna Pera, MD
Petra Redfors, MD
Jaume Roquer, MD, PhD
Jonathan Rosand, MD, MSc
Natalia S. Rost, MD, MPH
Peter M. Rothwell, MD

Author list continued on next page

ABSTRACT

Objective: The objective of this study was to assess the level of agreement between stroke subtype classifications made using the Trial of Org 10172 Acute Stroke Treatment (TOAST) and Causative Classification of Stroke (CCS) systems.

Methods: Study subjects included 13,596 adult men and women accrued from 20 US and European genetic research centers participating in the National Institute of Neurological Disorders and Stroke (NINDS) Stroke Genetics Network (SiGN). All cases had independently classified TOAST and CCS stroke subtypes. Kappa statistics were calculated for the 5 major ischemic stroke subtypes common to both systems.

Results: The overall agreement between TOAST and CCS was moderate (agreement rate, 70%; $\kappa = 0.59$, 95% confidence interval [CI] 0.58–0.60). Agreement varied widely across study sites, ranging from 28% to 90%. Agreement on specific subtypes was highest for large-artery atherosclerosis ($\kappa = 0.71$, 95% CI 0.69–0.73) and lowest for small-artery occlusion ($\kappa = 0.56$, 95% CI 0.54–0.58).

Conclusion: Agreement between TOAST and CCS diagnoses was moderate. Caution is warranted when comparing or combining results based on the 2 systems. Replication of study results, for example, genome-wide association studies, should utilize phenotypes determined by the same classification system, ideally applied in the same manner. *Neurology*® 2014;83:1653–1660

GLOSSARY

CI = confidence interval; **CCS** = Causative Classification of Stroke; **NINDS** = National Institute of Neurological Disorders and Stroke; **SiGN** = Stroke Genetics Network; **TOAST** = Trial of Org 10172 Acute Stroke Treatment.

Trial of Org 10172 in Acute Stroke Treatment (TOAST)[1] and the Causative Classification of Stroke (CCS)[2–4] are 2 well-established systems for classifying ischemic stroke. They use broadly similar categories of stroke diagnoses, e.g., large-vessel, small-vessel, and cardioembolic stroke, but may not necessarily be interchangeable. TOAST and CCS require different data and use different classification criteria and decision-making rules. It is therefore critical to understand the agreement rate between these 2 systems in diverse clinical and research settings. Delineation of the level of agreement between TOAST and CCS would be important to assess the validity of combining ischemic stroke subtyping using these 2 systems.

This report investigates the agreement between TOAST and CCS within the Stroke Genetics Network (SiGN), a collaborative study involving a network of international genetic research centers. This analysis is a retrospective pooled analysis of several independent research efforts, each of which enrolled patients under different research protocols.[5] TOAST and CCS were compared by assessing identical subtype assignment and accounting for agreement by chance by using a κ statistic. Because there is no gold standard in etiologic stroke classification, we make no qualitative judgments regarding which system is "better" at subtype assignment, rather report agreement to help inform whether the 2 systems make similar assignments.

Authors' affiliations are listed at the end of the article.

NINDS SiGN Study coinvestigators are listed on the *Neurology*® Web site at www.neurology.org.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Ralph L. Sacco, MD
Pankaj Sharma, MD
Agnieszka Slowik, MD
Cathie Sudlow, MD
Vincent Thijs, MD
Steffen Tiedt
Raffaella Valenti, MD
Bradford B. Worrall, MD, MSc
On behalf of the NINDS SiGN Study

Correspondence to Dr. Worrall: bbw9r@virginia.edu

Supplemental data at Neurology.org

METHODS The Stroke Genetics Network (SiGN) is a multinational collaboration with the goal of finding genetic determinants of stroke.[5] The SiGN Study standardized the phenotyping of the cases across all genetic research centers. The CCS system was chosen to facilitate study administration because of the Web-based, semiautomated, and evidence- and rule-based nature of the system (https://ccs.mgh.harvard.edu).[2] In addition to classifying stroke cases by subtype, the CCS system also has the practical benefit for large consortia of standardizing and centralizing all individual data points that underlie subtype classification.[2] A centralized committee of 4 expert neurologists met weekly to monitor subtype data quality and site performance. This panel aimed to ensure consistency of CCS assignments across all SiGN centers but did not contribute to subtype classifications directly.

CCS subtyping of stroke cases for this report was performed based on reviews of data available in study-specific case report forms and medical records by 41 adjudicators from 10 European and 10 US sites. Adjudicators included neurology residents (n = 10), neurologists (n = 17), stroke fellows (n = 12), one nurse, and one student. Adjudicators completed an interactive online training module and a certification module available at the CCS Web site (https://ccs.mgh.harvard.edu). Data adjudication began in June 2011 and was still ongoing at the time of data analysis for this report. This study included 13,596 cases adjudicated as of July 7, 2013. A centralized committee of 4 expert neurologists met weekly to monitor data quality and site performance. Feedback was provided during subtyping to ensure quality of data.

TOAST subtypes were determined locally by site investigators following individual study protocols without benefit of central oversight. Of note, TOAST subtypes were determined using the same data sources that were available for the CCS classifications. TOAST and CCS classifications were done by different physicians and at different time points in the majority of study sites but using the same study or site-specific case report forms. CCS adjudicators were required to confirm that they were fully blind to TOAST results before they began to enter patient data into CCS.

Two deviations from the above-mentioned protocol warrant acknowledgment. One center (STGEORGE) had completed case phenotyping using the CCS system before the initiation of the SiGN Study. Therefore, this center did not conduct CCS subtyping under the oversight of the expert panel. One other study (BASICMAR) utilized a computer algorithm rather than a certified adjudicator to extract data from a study data source to populate the required fields in CCS for 389 cases of their total of 1,088 cases.

For the purposes of this report, a complete investigation was defined as having head imaging (CT, MRI, or both), vascular imaging of both the intracranial and extracranial vasculature, and a cardiac evaluation consisting of echocardiography (either transthoracic or transesophageal) unless cardiac source of embolism was identified by medical history, physical examination, or ECG. Cranial imaging (either CT or MRI) was required for inclusion in the SiGN case set.

Statistical analysis was performed using SAS 9.2 (SAS Institute, Cary, NC). Percent of absolute agreement in subtype assignments is reported. Agreement was also estimated by κ statistics, and 95% confidence intervals (CIs) are provided for interpretation. Nonweighted κ values were calculated for 5 major stroke subtypes common to both systems: large-artery atherosclerosis, cardioaortic embolism, small-artery occlusion, other causes, and undetermined causes (i.e., cryptogenic causes, unclassifiable cases because of multiple competing etiologies, and incomplete evaluation).

RESULTS In total, 16,267 cases were enrolled in SiGN via the Web-based CCS system as of July 7, 2013. Of those, 13,596 stroke cases had subtypes classified using the TOAST system by the individual studies. Sample characteristics varied considerably by study design among the participating sites (table 1). For example, GEOS (Genetics of Early Onset Stroke)[6] targeted recruitment among young stroke patients. WHI, the Women's Health Initiative, recruited women only and reported low levels of current smoking. The diversity of study designs and populations allows for evaluation of the agreement between TOAST and CCS across a variety of clinical and research settings.

The overall agreement between CCS and TOAST was moderate (table 2) (κ = 0.59, 95% CI 0.58–0.60), although the agreement varied across study sites (χ^2 [df = 19] = 782; $p < 0.0001$) (table 2). Agreement on specific subtypes was highest for large-artery atherosclerosis (κ = 0.71) and lowest for small-artery occlusions (κ = 0.56). Table 3 provides the cross-tabulation for subtype agreement. The 2 systems identified approximately equal number of cases as undetermined (CCS 4,673 cases and TOAST 4,664 cases), but did not show much agreement on which those undetermined cases were (table 3). The agreement of TOAST and CCS for undetermined cases was only κ = 0.44 (95% CI 0.43–0.46). Agreement between TOAST and CCS regarding undetermined cases was primarily based on cases CCS determined to be "incomplete evaluations" (κ = 0.30). Cases classified by CCS as either "cryptogenic embolism" or "unclassified" had no agreement with the TOAST category of "undetermined" ($\kappa < 0.05$).

Stroke subtype agreement varied substantially across genetic research centers (see table 2). Part of this variability across study sites could be attributable to the variable process used to implement TOAST subtyping across sites. Additional variability could be attributable to the presence or absence of certain diagnostic evaluations available to each center (table 4). Agreement was slightly higher in the presence of vessel imaging, but slightly lower when a cardiac evaluation was performed. Agreement between TOAST and CCS was slightly lower when a complete evaluation was conducted (defined as the presence of brain imaging, cardiac evaluation, vascular imaging of the intra- and extracranial circulations). Regardless, the slight variation in agreement in the presence or absence of certain diagnostic evaluations (table 4) is not sufficient to account for the large variation seen across genetic research centers (table 2). The overall agreement reported here (κ = 0.59) belies the fact that in any particular center agreement ranged from excellent (STGEORGE κ = 0.85) to poor (BRAINS κ = 0.12).

Table 1 Descriptive characteristics of the cases in the SiGN Study with both CCS and TOAST classifications, by genetic research center

Study	No.	Age, y (SD)	% Female	% Vascular imaging	% Cardiac evaluation	% Head CT	% Brain MRI	% Complete investigation[a]
SiGN	13,596	66.6 (15.2)	48.1	65.1	79.2	92.5	57.0	54.8
BASICMAR	1,088	74.8 (11.7)	47.2	96.3	67.6	99.5	41.8	64.2
BRAINS	346	70.3 (13.8)	46.5	23.4	42.5	93.4	35.8	12.7
EDIN	620	71.0 (11.8)	45.5	1.3	46.3	79.0	25.5	0.7
GASROS	613	64.7 (14.9)	36.4	98.7	92.7	86.3	83.2	91.8
GCNKSS	642	67.3 (14.3)	50.2	54.8	84.6	93.3	58.3	47.8
GEOS	891	41.3 (6.9)	41.3	81.8	91.5	90.6	85.3	76.1
GRAZ	512	67.9 (14.3)	39.3	93.0	77.0	97.5	58.6	70.7
ISGS	675	63.6 (14.9)	43.1	91.4	80.4	91.0	84.3	73.9
KRAKOW	1,486	68.7 (14.0)	48.2	20.7	85.5	99.0	19.7	19.1
LEUVEN	524	67.6 (14.6)	41.8	91.0	96.9	92.4	84.9	88.4
MCISS	876	69.6 (14.7)	49.4	90.3	94.4	92.9	82.0	86.2
MIAMISR	314	62.6 (14.4)	35.0	97.8	99.4	99.0	85.7	97.1
MUNICH	524	66.7 (14.4)	40.8	99.8	92.4	86.8	82.3	92.4
NOMAS	438	69.3 (12.7)	54.3	77.9	95.9	98.2	48.2	77.4
OXVASC 2002-2008	554	74.2 (12.6)	50.9	23.6	56.7	93.5	24.4	15.0
REGARDS	489	71.7 (8.5)	46.8	52.4	74.8	90.6	65.4	43.6
SAHLSIS	1,083	55.6 (11.0)	35.5	50.5	79.0	98.7	58.1	41.1
STGEORGE	678	75.2 (12.9)	47.2	98.5	71.1	—	—	70.4
SWISS	407	62.9 (12.8)	46.7	67.1	70.0	64.6	47.7	51.4
WHI	836	74.0 (6.7)	100.0	36.8	73.2	89.2	55.5	28.6

Abbreviations: BASICMAR = BASe de datos de ICtus del hospital del MAR (Spain); BRAINS = Bio-Repository of DNA in Stroke (England); CCS = Causative Classification of Stroke; EDIN = Edinburgh stroke study (Scotland); GASROS = Genes Affecting Stroke Risk and Outcome Study (Boston); GCNKSS = Greater Cincinnati Northern Kentucky Stroke Study (Cincinnati); GEOS = Genetics of Early-Onset Stroke (Baltimore); GRAZ = Graz Stroke Study (Austria); ISGS = Ischemic Stroke Genetics Study (Jacksonville); KRAKOW = Krakow stroke study (Poland); LEUVEN = Leuven stroke study (Belgium); MCISS = Middlesex County Ischemic Stroke Study (New Jersey); MIAMISR = Miami Stroke Registry (Miami); MUNICH = Munich stroke study (Germany); NOMAS = Northern Manhattan Aging Study (New York); OXVASC = Oxford Vascular Study (England); REGARDS = Reasons for Geographic And Racial Differences in Stroke (Birmingham); SAHLSIS = Sahlgrenska Academy Study on Ischemic Stroke (Sweden); SiGN = Stroke Genetics Network; STGEORGE = St. George's Stroke Study (England); SWISS = Siblings With Ischemic Stroke Study (Jacksonville); TOAST = Trial of Org 10172 Acute Stroke Treatment; WHI = Women's Health Initiative (Boston).

[a] Complete investigation = head imaging (either CT or MRI or both), vascular imaging (requires both intracranial and extracranial arterial imaging), and cardiac evaluation (echocardiography performed unless cardiac source of embolism identified on physical examination and ECG).

A sensitivity analysis was performed removing the 2 centers that deviated from the network protocol. Removing the STGEORGE and relevant BASICMAR cases resulted in lower overall agreement ($\kappa = 0.57$).

DISCUSSION To accelerate advances in stroke treatment, prevention, and discovery of genetic and other novel risk factors, the heterogeneity of ischemic stroke must be addressed. Identifying the genetic determinants of many complex diseases has proven challenging[7]; stroke is no exception. Success is more likely to occur in large studies and active consortia of individual studies.[8] Standardization and harmonization of phenotypes will reduce misclassification error when combining analysis efforts in consortia. In the study of stroke, this often means the standardization of subtyping among cases.

Previously reported levels of agreement between the TOAST and CCS classification systems were high.[9,10] In a prospective cohort study of North Dublin, a single physician performed data abstraction and classification in both TOAST and CCS in 381 patients with first-ever ischemic stroke. An overall agreement was not reported, but agreement between the 2 systems on specific subtypes ranged from excellent ($\kappa = 0.95$ for cardioembolism) to moderate ($\kappa = 0.69$ for other and undetermined causes). Another study of 690 ischemic stroke patients from a single center (also included in this report, STGEORGE) reported excellent overall agreement ($\kappa = 0.85$). We report a lower overall agreement between the 2 systems, with some centers witnessing much less agreement between TOAST and CCS. This could be attributable to differences among the studies in their ability to take into account the whole spectrum

Table 2 Agreement statistics, κ (95% confidence interval), between CCS and TOAST for the SiGN Study

Study	No.	% Agreed	Overall κ	CE κ	LAA κ	SAO κ	Other κ	Undetermined κ
SiGN	13,596	0.70	0.59 (0.58-0.60)	0.68 (0.67-0.70)	0.71 (0.69-0.73)	0.56 (0.54-0.58)	0.64 (0.61-0.67)	0.44 (0.43-0.46)
BASICMAR	1,088	0.87	0.81 (0.78-0.84)	0.85 (0.82-0.88)	0.82 (0.78-0.87)	0.89 (0.86-0.92)	—	—
BRAINS	346	0.28	0.12 (0.07-0.17)	0.30 (0.19-0.42)	0.14 (0.05-0.24)	0.09 (0.02-0.17)	—	0.03 (−0.06 to 0.11)
EDIN	620	0.69	0.47 (0.41-0.53)	0.64 (0.55-0.72)	0.76 (0.67-0.84)	0.31 (0.21-0.40)	—	0.36 (0.29-0.43)
GASROS	613	0.69	0.59 (0.54-0.64)	0.62 (0.55-0.68)	0.69 (0.62-0.76)	0.57 (0.46-0.67)	0.67 (0.58-0.77)	0.42 (0.33-0.51)
GCNKSS	642	0.78	0.69 (0.65-0.74)	0.78 (0.72-0.84)	0.81 (0.74-0.87)	0.65 (0.58-0.73)	0.70 (0.45-0.95)	0.59 (0.52-0.65)
GEOS	891	0.66	0.53 (0.49-0.58)	0.61 (0.55-0.68)	0.70 (0.62-0.78)	0.69 (0.63-0.76)	0.42 (0.34-0.51)	0.39 (0.33-0.44)
GRAZ	512	0.78	0.70 (0.66-0.75)	0.74 (0.68-0.80)	0.87 (0.81-0.92)	0.64 (0.54-0.74)	0.84 (0.72-0.96)	0.54 (0.45-0.62)
ISGS	675	0.63	0.51 (0.46-0.56)	0.55 (0.48-0.62)	0.64 (0.57-0.71)	0.40 (0.31-0.50)	0.78 (0.67-0.90)	0.36 (0.29-0.43)
KRAKOW	1,486	0.75	0.62 (0.59-0.65)	0.67 (0.63-0.71)	0.83 (0.79-0.87)	0.22 (0.12-0.32)	0.83 (0.73-0.92)	0.52 (0.47-0.56)
LEUVEN	524	0.67	0.54 (0.49-0.60)	0.62 (0.55-0.68)	0.67 (0.58-0.75)	0.49 (0.36-0.62)	0.79 (0.67-0.91)	0.33 (0.25-0.42)
MCISS	876	0.61	0.50 (0.46-0.54)	0.56 (0.50-0.62)	0.57 (0.51-0.63)	0.74 (0.67-0.81)	0.60 (0.49-0.71)	0.22 (0.16-0.28)
MIAMISR	314	0.68	0.58 (0.52-0.65)	0.66 (0.57-0.74)	0.63 (0.53-0.73)	0.68 (0.58-0.79)	0.48 (0.31-0.65)	0.34 (0.20-0.48)
MUNICH	524	0.63	0.51 (0.45-0.56)	0.65 (0.58-0.72)	0.53 (0.45-0.62)	0.42 (0.27-0.57)	0.67 (0.55-0.79)	0.30 (0.22-0.39)
NOMAS	438	0.65	0.54 (0.48-0.60)	0.55 (0.47-0.64)	0.76 (0.67-0.84)	0.65 (0.56-0.73)	0.36 (0.00-0.72)	0.31 (0.22-0.40)
OXVASC 2002-2008	554	0.76	0.66 (0.61-0.71)	0.83 (0.78-0.88)	0.88 (0.81-0.95)	0.39 (0.30-0.49)	1.00 (1.00-1.00)	0.57 (0.50-0.63)
REGARDS	489	0.62	0.47 (0.41-0.53)	0.47 (0.37-0.56)	0.65 (0.56-0.75)	0.45 (0.34-0.56)	0.56 (0.41-0.72)	0.37 (0.28-0.45)
SAHLSIS	1,083	0.68	0.56 (0.52-0.60)	0.70 (0.65-0.76)	0.73 (0.66-0.79)	0.31 (0.23-0.38)	0.84 (0.79-0.89)	0.41 (0.35-0.46)
STGEORGE	678	0.90	0.85 (0.82-0.89)	0.91 (0.87-0.94)	0.89 (0.85-0.94)	0.85 (0.79-0.90)	—	0.78 (0.72-0.83)
SWISS	407	0.54	0.38 (0.32-0.44)	0.45 (0.33-0.56)	0.65 (0.56-0.74)	0.13 (0.06-0.20)	0.53 (0.34-0.72)	0.27 (0.18-0.36)
WHI	836	0.64	0.50 (0.45-0.54)	0.66 (0.59-0.72)	0.61 (0.52-0.70)	0.46 (0.39-0.53)	0.45 (0.30-0.60)	0.38 (0.32-0.45)

Abbreviations: BASICMAR = BASe de datos de ICtus del hospital del MAR (Spain); BRAINS = Bio-Repository of DNA in Stroke (England); CCS = Causative Classification of Stroke; CE = cardiac embolism; EDIN = Edinburgh stroke study (Scotland); GASROS = Genes Affecting Stroke Risk and Outcome Study (Boston); GCNKSS = Greater Cincinnati Northern Kentucky Stroke Study (Cincinnati); GEOS = Genetics of Early-Onset Stroke (Baltimore); GRAZ = Graz Stroke Study (Austria); ISGS = Ischemic Stroke Genetics Study (Jacksonville); KRAKOW = Krakow stroke study (Poland); LAA = large-artery atherosclerosis; LEUVEN = Leuven stroke study (Belgium); MCISS = Middlesex County Ischemic Stroke Study (New Jersey); MIAMISR = Miami Stroke Registry (Miami); MUNICH = Munich stroke study (Germany); NOMAS = Northern Manhattan Aging Study (New York); OXVASC = Oxford Vascular Study (England); REGARDS = Reasons for Geographic And Racial Differences in Stroke (Birmingham); SAHLSIS = Sahlgrenska Academy Study on Ischemic Stroke (Sweden); SAO = small-artery occlusion; SiGN = Stroke Genetics Network; STGEORGE = St. George's Stroke Study (England); SWISS = Siblings With Ischemic Stroke Study (Jacksonville); TOAST = Trial of Org 10172 Acute Stroke Treatment; WHI = Women's Health Initiative (Boston).
Subtype-specific kappas are missing for some studies because of either one classification system, or both, not classifying any cases as the relevant subtype.

Table 3 Cross-classification of 13,596 cases by TOAST and CCS

	TOAST					
	LAA	CE	SAO	Other	Undetermined	Total
CCS						
LAA	1,691	113	83	18	453	2,358
CE	103	3,063	155	27	761	4,109
SAO	27	68	1,297	12	296	1,700
Other	42	59	22	439	203	765
Undetermined	246	485	902	71	2,960	4,664
Total	2,109	3,788	2,459	567	4,673	13,596

Abbreviations: CCS = Causative Classification of Stroke; CE = cardiac embolism; LAA = large-artery atherosclerosis; SAO = small-artery occlusion; TOAST = Trial of Org 10172 Acute Stroke Treatment.

of variance in determining etiologic stroke subtypes. The present study, in our opinion, offers less bias because of its larger sample size, higher number of raters, multicenter design, and methodology that required a blinded assessment of CCS and TOAST subtypes. The current report provides agreement statistics stratified by study center and demonstrates variability in agreement rates with many genetic research centers having poor agreement (14 of 20 have $\kappa < 0.60$). One could interpret our results to indicate that in many research settings, the agreement between the 2 systems is quite low and data from standardized chart abstractions studies may not reflect the complexities of many practical implementations.

The lack of good agreement between TOAST and CCS is not surprising because these 2 systems use different classification criteria, definitions for subtypes, and diagnostic investigation requirements (table 5). Furthermore, their internal reliability also differs; existing studies by independent investigators demonstrate a moderate interrater reliability for the TOAST classification system with κ values ranging between 0.42 and 0.54.[11–15] In contrast, interrater reliability of the CCS ranges between 0.8 and 0.9.[2–4] These studies tended to be small and with varying methodologies. Well-powered reliability studies of both systems are still needed. While TOAST and CCS differ from each other in several ways, they both are subject to variability because of differences in adjudicators' ability to

Table 4 Agreement statistics, κ (95% confidence interval), between CCS and TOAST for the SiGN Study by the presence of diagnostic evaluations

	No.	Overall κ	CE κ	LAA κ	SAO κ	Other κ	Undetermined κ
Vascular imaging							
Present	8,846	0.61 (0.60-0.62)	0.67 (0.65-0.68)	0.70 (0.68-0.72)	0.67 (0.65-0.69)	0.67 (0.64-0.70)	0.40 (0.37-0.42)
Absent	4,750	0.54 (0.52-0.56)	0.72 (0.70-0.74)	0.69 (0.65-0.73)	0.34 (0.30-0.37)	0.51 (0.43-0.59)	0.45 (0.43-0.48)
Cardiac evaluation							
Present	10,768	0.57 (0.56-0.58)	0.66 (0.64-0.67)	0.68 (0.66-0.70)	0.56 (0.54-0.59)	0.61 (0.58-0.65)	0.40 (0.38-0.42)
Absent	2,828	0.61 (0.59-0.64)	0.35 (0.23-0.46)	0.80 (0.77-0.83)	0.53 (0.50-0.57)	0.74 (0.69-0.80)	0.54 (0.51-0.57)
Head CT							
Present	11,948	0.58 (0.57-0.59)	0.67 (0.66-0.69)	0.70 (0.68-0.72)	0.55 (0.53-0.57)	0.65 (0.62-0.68)	0.43 (0.41-0.45)
Absent	970	0.51 (0.47-0.55)	0.59 (0.53-0.65)	0.68 (0.61-0.74)	0.47 (0.40-0.54)	0.58 (0.47-0.69)	0.37 (0.31-0.43)
Brain MRI							
Present	7,358	0.57 (0.56-0.59)	0.62 (0.60-0.65)	0.69 (0.66-0.71)	0.61 (0.59-0.63)	0.66 (0.63-0.70)	0.40 (0.37-0.42)
Absent	5,560	0.58 (0.56-0.60)	0.71 (0.69-0.73)	0.72 (0.69-0.75)	0.43 (0.39-0.46)	0.55 (0.49-0.62)	0.46 (0.44-0.49)
Complete evaluation							
Present	7,451	0.58 (0.56-0.59)	0.65 (0.63-0.66)	0.67 (0.65-0.69)	0.64 (0.61-0.67)	0.63 (0.59-0.67)	0.37 (0.34-0.39)
Absent	6,145	0.60 (0.58-0.61)	0.73 (0.71-0.75)	0.77 (0.75-0.80)	0.48 (0.45-0.51)	0.66 (0.61-0.71)	0.49 (0.47-0.51)

Abbreviations: CCS = Causative Classification of Stroke; CE = cardiac embolism; LAA = large-artery atherosclerosis; SAO = small-artery occlusion; SiGN = Stroke Genetics Network; TOAST = Trial of Org 10172 Acute Stroke Treatment.

Table 5 Characteristics of the TOAST and CCS classification systems

	TOAST	CCS
Year of publication	1993	2005
Diagnosis of LAA	Requires imaging of a limited portion of the extracranial circulation	Result influenced by intracranial imaging (if performed)
Diagnosis of SAO	Does not require imaging confirmation	Does require imaging confirmation
Size limit for lacunar infarct	15 mm	20 mm
Imaging of the parent artery in lacunar infarcts required	No	Yes
Threshold to separate high- and low-risk cardiac sources	No	2% absolute primary risk threshold
Criteria to identify the most likely etiology in the presence of multiple etiologies	No	Yes
Criteria to identify a known subtype in patients with missing tests	No	Yes

Abbreviations: CCS = Causative Classification of Stroke; LAA = large-artery atherosclerosis; SAO = small-artery occlusion; TOAST = Trial of Org 10172 Acute Stroke Treatment.

interpret diagnostic test findings as well as variability in the completeness and quality of available diagnostic investigations. Regarding the latter point, we found that agreement for a subtype was generally higher when diagnostic investigations were complete for that particular subtype. For instance, in patients with complete cardiac evaluation, the agreement for cardiac embolism was almost twice as high compared to those with incomplete cardiac investigation ($\kappa = 0.66$ vs 0.35; table 4). Likewise, agreement for small-artery occlusion was higher in the presence of brain MRI ($\kappa = 0.61$ vs 0.43). In contrast to cardiac embolism and small-artery occlusion, there was no difference in agreement for large-artery atherosclerosis between cases with and without complete vascular evaluation. This may be attributable to diagnosis of large-artery atherosclerosis being contingent on extracranial carotid artery stenosis—the most common site for large-artery atherosclerosis—which does not require a complete assessment of both extracranial and intracranial vessels.

These findings suggest that availability of objective diagnostic information reduces the subjective component in decision-making for stroke subtypes regardless of the classification system used. Nevertheless, too much diagnostic information provides the opportunity for differential investigator interpretation in the absence of rule-based criteria and this may in turn reduce agreement between CCS and TOAST. In line with this, we found that overall agreement rate was lower, albeit slightly, when all investigations (brain imaging, brain vascular imaging, and cardiac evaluation) were complete than when they were incomplete ($\kappa = 0.58$ vs 0.60). We also found that the agreement rate for the undetermined group was lower than rates in other etiologic categories. The undetermined group is a heterogeneous category consisting of cryptogenic stroke (undetermined-unknown), multiple competing etiologies (undetermined-unclassified), and missing diagnostic tests (incomplete evaluation). Lower agreement rate in the undetermined category reflects differences between TOAST and CCS in dealing with multiple potential causes and missing diagnostic tests. CCS takes into account the completeness of diagnostic investigations and strength of evidence favoring one mechanism over others in the presence of multiple mechanisms in identifying stroke subtypes. In contrast, TOAST provides limited guidance on these issues leading to room for opinion in many practical implementations and hence variance in subtype assignments.

The present study required a uniform Web-based training and certification of investigators to be able to perform CCS. The same standardization was not applied to the TOAST classification. The TOAST classification was done locally, using local interpretations of the TOAST classification system, and before the formation of the SiGN collaboration. This differential application of TOAST is likely responsible for the variability in agreement seen between the centers. Thus, the overall agreement captures both differences in the subtyping systems and differences in their applications. The optimal test to compare the 2 classifications would have included prospective data-quality assessment and centrally trained certified investigators for performing also the TOAST classification at the same time as the CCS. However, we address this limitation by assessing agreement separately within each genetic research center, and the agreement between the 2 systems was modest at best for a majority of them. In only 2 of the 20 studies can agreement between TOAST and CCS be classified as excellent ($\kappa > 0.80$). Of note, both of those studies used protocols for CCS assignment that deviated from the

recommended consortium design. In addition, a computerized tool for TOAST classification has also been made available[12] and an additional question of interest may be how well the computerized TOAST agrees with CCS.

The agreement between CCS and TOAST reported here is lower than previously reported. The low agreement between the 2 systems described here simply means that the 2 systems classify stroke cases in different categories, although perhaps unfortunately, the names of the categories are similar. The practical implication of this finding is that combining or comparing classifications across systems should proceed with caution, and where possible, rephenotyping should be encouraged before combining data. For example, replication of results from genetic association studies should be made using phenotypes from the same classification system. A large benefit of the CCS system is the standardization of input and output data across cases from different sites. This feature allows for flexible analysis and further stratification of stroke phenotypes and hence promises utility in genetic studies such as SiGN.

AUTHOR AFFILIATIONS

From the University of Maryland School of Medicine (P.F.M.), Division of Endocrinology, Diabetes and Nutrition; University of Maryland (S.J.K., J.W.C., M.H.); Massachusetts General Hospital (H.A.), Departments of Neurology and Radiology, Harvard Medical School; Mayo Clinic (R.D.B.), Rochester, MN; Mayo Clinic Florida (J.F.M.), Jacksonville; University of Miami (T.R., M.K., R.L.S.); Albert Einstein College of Medicine (S.W.-S., R.C.K., D.L.L.); University of Cincinnati (D.W.); Skåne University Hospital (G.A., H.D.); Massachusetts General Hospital and Broad Institute (A.B., J. Rosand, N.S.R.); University of Alabama Birmingham (D.A.B.); NINDS-NIH (R.C., K.G.); University Medical Center Utrecht (P.I.W.d.B.), the Netherlands; Institute for Stroke and Dementia Research (M.D., S.T.), Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich; Munich Cluster for Systems Neurology (SyNergy) (M.D.), Germany; Neuroscience Institute (R.P.G., L.R.P.), Saint Francis Medical Center; Sahlgrenska Academy at University of Gothenburg (C.J., K.J., A.P., P.R.); IMIM-Hospital Universitari del Mar (J.J.-C., J. Roquer); Medical University Graz (P.K.); Laboratory of Neurobiology (R.L., V.T.), Vesalius Research Center, VIB, Leuven; Experimental Neurology (Department of Neurosciences) and Leuven Research Institute for Neuroscience and Disease (LIND) (R.L., V.T.), University of Leuven (KU Leuven); Department of Neurology (R.L., V.T.), University Hospitals Leuven, Belgium; Oxford University (L.L.); Lund University and Skåne University Hospital (A.L.); St. George's University of London (H.S.M., R.V.); Jagiellonian University Medical College (J.P.); Nuffield Department of Clinical Neurosciences (P.M.R.), University of Oxford; John Radcliffe Hospital (P.M.R.), Oxford, UK; Imperial College London (P.S.); Jagiellonian University Medical College (A.S.); University of Edinburgh (C.S.); and University of Virginia Health System (B.B.W.), Departments of Neurology and Health Evaluation Sciences.

AUTHOR CONTRIBUTIONS

Patrick F. McArdle: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data, statistical analysis, study supervision. Steven J. Kittner: drafting/revising the manuscript, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, study supervision, obtaining funding. Hakan Ay: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data, statistical analysis, study supervision. Robert D. Brown, Jr.: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval. James F. Meschia: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data, study supervision. Tatjana Rundek: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, critical revision of the manuscript. Sylvia Wassertheil-Smoller: drafting/revising the manuscript, study concept or design, accepts responsibility for conduct of research and will give final approval, acquisition of data, obtaining funding. Daniel Woo: drafting/revising the manuscript, study concept or design, accepts responsibility for conduct of research and will give final approval, acquisition of data, study supervision, obtaining funding. Gunnar Andsberg: drafting/revising the manuscript, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval. Alessandro Biffi: analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval. David A. Brenner: analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, medical chart review of more than 600 stroke patients, computer entry into the CCS system. John W. Cole: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval. Roderick Corriveau: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, study supervision. Paul I.W. de Bakker: study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data. Hossein Delavaran: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Martin Dichgans: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Raji P. Grewal: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data, study supervision. Katrina Gwinn: drafting/revising the manuscript, study concept or design, accepts responsibility for conduct of research and will give final approval, study supervision. Mohammed Huq: analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data. Christina Jern: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Jordi Jimenez-Conde: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Katarina Jood: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Robert C. Kaplan: study concept or design, accepts responsibility for conduct of research and will give final approval, acquisition of data. Petra Katschnig: analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval. Michael Katsnelson: analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data. Daniel L. Labovitz: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Robin Lemmens: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Linxin Li: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Arne Lindgren: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data, obtaining funding. Hugh S. Markus: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Leema R. Peddareddygari: analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data, CCS adjudicator. Annie Pedersén: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Joanna Pera: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Petra Redfors: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data. Jaume Roquer:

analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data. Jonathan Rosand: drafting/revising the manuscript, study concept or design, accepts responsibility for conduct of research and will give final approval, acquisition of data, obtaining funding. Natalia S. Rost: drafting/revising the manuscript, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data. Peter M. Rothwell: study concept or design, accepts responsibility for conduct of research and will give final approval, contribution of vital reagents/tools/patients, acquisition of data. Ralph L. Sacco: drafting/revising the manuscript, accepts responsibility for conduct of research and will give final approval, acquisition of data, study supervision. Pankaj Sharma: drafting/revising the manuscript, study concept or design, accepts responsibility for conduct of research and will give final approval, acquisition of data. Agnieszka Slowik: analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data. Cathie Sudlow: drafting/revising the manuscript, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data. Vincent Thijs: drafting/revising the manuscript, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data. Steffen Tiedt: analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data. Raffaella Valenti: analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data. Bradford B. Worrall: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, accepts responsibility for conduct of research and will give final approval, acquisition of data, study supervision, obtaining funding. The writing group included Dr. McArdle, Dr. Kittner, Dr. Ay, Dr. Brown, Dr. Meschia, Dr. Rundek, Dr. Wassertheil-Smoller, Dr. Woo, and Dr. Worrall.

STUDY FUNDING

The SiGN Network was funded by a cooperative agreement grant from the National Institute of Neurological Disorders and Stroke (NINDS): U01 NS069208. Funding for network members is presented in appendix e-1 on the *Neurology*® Web site at Neurology.org.

DISCLOSURE

The authors report no disclosures relevant to the manuscript. Go to Neurology.org for full disclosures.

Received March 19, 2014. Accepted in final form August 4, 2014.

REFERENCES

1. Adams HP Jr, Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke: definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in acute stroke treatment. Stroke 1993;24:35–41.
2. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: the causative classification of stroke system. Stroke 2007;38:2979–2984.
3. Ay H, Furie KL, Singhal A, Smith WS, Sorensen AG, Koroshetz WJ. An evidence-based causative classification system for acute ischemic stroke. Ann Neurol 2005;58:688–697.
4. Arsava EM, Ballabio E, Benner T, et al. The causative classification of stroke system: an international reliability and optimization study. Neurology 2010;75:1277–1284.
5. Meschia JF, Arnett DK, Ay H, et al. Stroke Genetics Network (SiGN) Study: design and rationale for a genome-wide association study of ischemic stroke subtypes. Stroke 2013;44:2694–2702.
6. Kittner SJ, Stern BJ, Wozniak M, et al. Cerebral infarction in young adults: the Baltimore–Washington Cooperative Young Stroke Study. Neurology 1998;50:890–894.
7. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet 2012;90:7–24.
8. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747–753.
9. Lanfranconi S, Markus HS. Stroke subtyping for genetic association studies? A comparison of the CCS and TOAST classifications. Int J Stroke 2012;8:626–631.
10. Marnane M, Duggan CA, Sheehan OC, et al. Stroke subtype classification to mechanism-specific and undetermined categories by TOAST, A-S-C-O, and causative classification system: direct comparison in the North Dublin Population Stroke Study. Stroke 2010;41:1579–1586.
11. Gordon DL, Bendixen BH, Adams HP Jr, Clarke W, Kappelle LJ, Woolson RF. Interphysician agreement in the diagnosis of subtypes of acute ischemic stroke: implications for clinical trials. The TOAST Investigators. Neurology 1993;43:1021–1027.
12. Goldstein LB, Jones MR, Matchar DB, et al. Improving the reliability of stroke subgroup classification using the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) criteria. Stroke 2001;32:1091–1098.
13. Atiya M, Kurth T, Berger K, Buring JE, Kase CS. Interobserver agreement in the classification of stroke in the Women's Health Study. Stroke 2003;34:565–567.
14. Meschia JF, Barrett KM, Chukwudelunzu F, et al. Interobserver agreement in the Trial of Org 10172 in Acute Stroke Treatment classification of stroke based on retrospective medical record review. J Stroke Cerebrovasc Dis 2006;15:266–272.
15. Selvarajah JR, Glaves M, Wainwright J, Jha A, Vail A, Tyrrell PJ. Classification of minor stroke: intra- and interobserver reliability. Cerebrovasc Dis 2009;27:209–214.

Agreement between TOAST and CCS ischemic stroke classification: The NINDS SiGN Study

Patrick F. McArdle, Steven J. Kittner, Hakan Ay, et al.

Neurology 2014;83;1653-1660 Published Online before print September 26, 2014
DOI 10.1212/WNL.0000000000000942

4.3. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study

The identification of disease-associated genetic variants represents an important strategy for the discovery of disease-causing molecular pathways and eventually the development of new diagnostic tests and therapies. Genome-wide association studies (GWAS) serve as a tool to link these single-nucleotide polymorphisms (SNP) with a phenotype, but have been limited by small sample sizes and phenotypic heterogeneity in the context of IS.

In a first step, 16.851 IS cases and 32.473 controls from different ancestries were enrolled for genome-wide analysis. 1372 SNPs in 268 loci linked to IS or any etiological subgroup were identified. The second step included an additional set of 20.941 IS cases and 364.736 controls from 18 independent studies enabling *in silico* analysis of overall 37.893 IS cases and 397.209 controls of the SNPs identified in the first step. By this, we identified a novel association of large artery atherosclerosis-related stroke with a locus near *TSPAN2* (joint odds ratio 1.19, 95 % CI 1.12 – 1.26, $p=1.3x10^{-9}$). Four previously established loci, *PITX2* and *ZFHX3* for cardioembolic stroke, *HDAC9* for large artery atherosclerosis-related stroke, and 12q24 near *ALDH2* with small-artery stroke, were confirmed.

In addition to the identification of a novel locus associated with large artery atherosclerosis-related stroke, these findings indicate the subtype-specificity of all IS-related loci implying different genetic signatures of IS subtypes.

Author contribution: recruitment of IS patients, collection of blood samples, classification of IS patients according to their etiology (please see section 10 for further details).

Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study

*NINDS Stroke Genetics Network (SiGN) and International Stroke Genetics Consortium (ISGC)**

Lancet Neurol 2016; 15: 174–84

Published Online December 18, 2015 http://dx.doi.org/10.1016/S1474-4422(15)00338-5

This online publication has been corrected. The corrected version first appeared at thelancet.com/neurology on January 15, 2016

See Comment page 130

*Authors listed at end of the paper

Correspondence to: Dr Jonathan Rosand, Department of Neurology and Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA
jrosand@partners.org

Summary

Background The discovery of disease-associated loci through genome-wide association studies (GWAS) is the leading genetic approach to the identification of novel biological pathways underlying diseases in humans. Until recently, GWAS in ischaemic stroke have been limited by small sample sizes and have yielded few loci associated with ischaemic stroke. We did a large-scale GWAS to identify additional susceptibility genes for stroke and its subtypes.

Methods To identify genetic loci associated with ischaemic stroke, we did a two-stage GWAS. In the first stage, we included 16 851 cases with state-of-the-art phenotyping data and 32 473 stroke-free controls. Cases were aged 16 to 104 years, recruited between 1989 and 2012, and subtypes of ischaemic stroke were recorded by centrally trained and certified investigators who used the web-based protocol, Causative Classification of Stroke (CCS). We constructed case-control strata by identifying samples that were genotyped on nearly identical arrays and were of similar genetic ancestral background. We cleaned and imputed data by use of dense imputation reference panels generated from whole-genome sequence data. We did genome-wide testing to identify stroke-associated loci within each stratum for each available phenotype, and we combined summary-level results using inverse variance-weighted fixed-effects meta-analysis. In the second stage, we did in-silico lookups of 1372 single nucleotide polymorphisms identified from the first stage GWAS in 20 941 cases and 364 736 unique stroke-free controls. The ischaemic stroke subtypes of these cases had previously been established with the Trial of Org 10 172 in Acute Stroke Treatment (TOAST) classification system, in accordance with local standards. Results from the two stages were then jointly analysed in a final meta-analysis.

Findings We identified a novel locus (G allele at rs12122341) at 1p13.2 near *TSPAN2* that was associated with large artery atherosclerosis-related stroke (first stage odds ratio [OR] 1·21, 95% CI 1·13–1·30, $p=4·50\times10^{-8}$; joint OR 1·19, 1·12–1·26, $p=1·30\times10^{-9}$). Our results also supported robust associations with ischaemic stroke for four other loci that have been reported in previous studies, including *PITX2* (first stage OR 1·39, 1·29–1·49, $p=3·26\times10^{-19}$; joint OR 1·37, 1·30–1·45, $p=2·79\times10^{-32}$) and *ZFHX3* (first stage OR 1·19, 1·11–1·27, $p=2·93\times10^{-7}$; joint OR 1·17, 1·11–1·23, $p=2·29\times10^{-10}$) for cardioembolic stroke, and *HDAC9* (first stage OR 1·29, 1·18–1·42, $p=3·50\times10^{-8}$; joint OR 1·24, 1·15–1·33, $p=4·52\times10^{-9}$) for large artery atherosclerosis stroke. The 12q24 locus near *ALDH2*, which has previously been associated with all ischaemic stroke but not with any specific subtype, exceeded genome-wide significance in the meta-analysis of small artery stroke (first stage OR 1·20, 1·12–1·28, $p=6·82\times10^{-8}$; joint OR 1·17, 1·11–1·23, $p=2·92\times10^{-9}$). Other loci associated with stroke in previous studies, including *NINJ2*, were not confirmed.

Interpretation Our results suggest that all ischaemic stroke-related loci previously implicated by GWAS are subtype specific. We identified a novel gene associated with large artery atherosclerosis stroke susceptibility. Follow-up studies will be necessary to establish whether the locus near *TSPAN2* can be a target for a novel therapeutic approach to stroke prevention. In view of the subtype-specificity of the associations detected, the rich phenotyping data available in the Stroke Genetics Network (SiGN) are likely to be crucial for further genetic discoveries related to ischaemic stroke.

Funding US National Institute of Neurological Disorders and Stroke, National Institutes of Health.

Introduction

Worldwide, stroke is the second leading cause of death[1] and a major contributor to dementia and age-related cognitive decline. About 15 million people have a stroke each year.[1] Most survivors are left with a permanent disability, which makes stroke the world's leading cause of adult incapacity.[2] Strokes result from the sudden occlusion or rupture of a blood vessel supplying the brain, and so are categorised accordingly as ischaemic (vessel occlusion) or haemorrhagic (vessel rupture) on the basis of neuroimaging results. Up to 85% of all strokes are ischaemic.

Although hypertension, atrial fibrillation, diabetes mellitus, and cigarette smoking are known risk factors for stroke,[3] a substantial proportion of the risk remains unexplained and might be attributable to inherited genetic variation. Discovery of genetic variants that predispose to stroke is a crucial first step toward the development of improved diagnostic tests for stroke and novel therapies that might reduce the disease burden. Genome-wide association studies (GWAS) have thus far identified only a few confirmed loci,[4–7] which together account for a small proportion of the heritable risk.[8]

Ischaemic stroke occurs when the blood flow to a region of the brain is interrupted because of blockage of a blood vessel. Because vessel occlusion can occur through different mechanisms, ischaemic stroke can be subtyped

Research in context

Evidence before this study

We searched PubMed with the search terms "stroke" and "genome wide association study" for reports published before Oct 19, 2015. We only included peer-reviewed reports in English. Compared with the rapid pace of genetic discovery for other common disorders, only four loci (PITX2, HDAC9, ZFHX3, and 12q24.2) have been convincingly implicated by genome-wide association studies (GWAS) in ischaemic stroke. GWAS of stroke have been limited by small sample sizes and concerns about phenotypic heterogeneity.

Added value of this study

To our knowledge, the National Institute of Neurological Disorders and Stroke (NINDS)-Stroke Genetics Network (SiGN) project is the largest and most comprehensive study of ischaemic stroke so far. Discovery analyses were done in 16 851 cases and 32 473 controls and findings were followed up in an additional 20 941 cases and 364 736 controls. Furthermore, the project implemented the Causative Classification of Stroke (CCS) system to subtype cases and generated a rich phenotypic database. Trial of Org 10 172 in Acute Stroke Treatment (TOAST)-based subtypes were also available, allowing for the first ever analysis of the genetic overlap between TOAST and CCS subtypes.

Implications of all the available evidence

Our data show that increasing sample size and applying a standardised subtyping method can reveal additional information about the underlying genetic architecture of stroke. Because we had access to phenotype information generated by two different subtyping methods, we also showed that there is moderate to strong genetic correlation between the CCS and TOAST subtyping methods, suggesting that future studies might benefit from liberal inclusion of cases, regardless of subtyping approach. Also, our results show that all discovered loci, including the 12q24.12 locus, which was previously implicated in all ischaemic stroke, are specific to a single subtype, suggesting that these subtypes will have at least partly distinct genetic signatures. Because of the subtype-specificity of genetic associations in stroke, substantially larger samples of stroke subtypes will probably be needed to expand the number of identified stroke loci to that of other common diseases.

on the basis of the presumed mechanism: large artery atherosclerosis, cardioembolism, or small artery occlusion. With one exception, all associations for ischaemic stroke detected in GWAS have been subtype-specific, suggesting the need for studies that are powered to detect subtype-specific associations. The National Institute of Neurological Disorders and Stroke (NINDS) Stroke Genetics Network (NINDS-SiGN)[9] is the largest and most comprehensive GWAS of stroke and its subtypes to date. We sought to detect new associations of polymorphisms with risk of ischaemic stroke and its subtypes and to provide evidence for previously reported associations.

Methods

Study design

We did a two-stage joint association analysis of ischaemic stroke and its subtypes. The first stage consisted of a GWAS, and the second stage was an in-silico association analysis of the top single nucleotide polymorphisms (SNPs) identified in the first stage in a set of independent samples of cases and controls. We then analysed both stages together to identify loci that exceeded the threshold for genome-wide significance (1×10^{-8}). Compared with separate discovery and replication analyses, this two-stage approach has been shown to improve the power for discovery without altering the type I error.[10]

Study sample

For the first stage, we assessed 31 existing collections that included cases of ischaemic stroke with either available genotypic data or DNA for genotyping, neuroimaging confirmation of stroke, and adequate clinical data to enable phenotypic classification. The cases of ischaemic stroke in the second stage met similar requirements, except that we used pre-existing Trial of Org 10 172 in Acute Stroke Treatment (TOAST)[11] subtyping data for the phenotypic classification. The appendix contains details about each collection, including their study design.

See Online for appendix

For each collection, approval for inclusion in the SiGN analysis complied with local ethical standards and with local institutional review board and ethics committee oversight. All people included as cases and controls provided written informed consent for genetic studies either directly or by a legally authorised representative.

Classification of stroke subtype

In the NINDS-SiGN,[9] we used two subtyping systems: the Causative Classification of Stroke (CCS) system, which is a standardised web-based subtype classification system,[12] and the more widely used TOAST subtype classification system.[11,13] Both of these systems are based on a similar conceptual framework but are operationalised differently. The TOAST subtyping system is based on the application of written rules requiring clinician judgment; patients with conflicting potential causes are placed into an undetermined category. The CCS subtyping system uses two web-based algorithms that classify patients with conflicting potential causes. Causative (CCSc) categorisation uses historical examination and test data from each ischaemic stroke subject to assign the most probable cause in the presence of competing aetiologies, while phenotypic (CCSp) categorisation uses abnormal test findings to assign each case into one or more major groups without using rules to determine the most likely aetiology. In addition to the generation of both CCSc and CCSp subtype categories, the advantages of the CCS

	Location of sample collection	Genotyping platform	Ancestry groups	Cases	Controls
First stage					
Case-control group 1					
BRAINS	UK	650Q	European	267	..
MGH-GASROS	USA	610	European	111	..
ISGS	USA	610	European	351	..
SWISS	USA	610	European	25	..
HABC	USA	1M	European	..	1586
Case-control group 2					
EDIN	UK	660	European	566	..
MUNICH	UK	660	European	1131	..
OXVASC	UK	660	European	457	..
STGEORGE	UK	660	European	418	..
KORA	Germany	550	European	..	804
WTCCC	UK	660	European	..	5150
Case-control group 3					
GEOS	USA	1M	African, European	843	880
Case-control group 4					
BRAINS	UK	5M	European, Hispanic	110	..
MGH-GASROS	USA	5M	African, European, Hispanic	456	..
GCNKSS	USA	5M	African, European, Hispanic	482	..
ISGS	USA	5M	African, European, Hispanic	178	..
MCISS	USA	5M	African, European, Hispanic	619	..
MIAMISR	USA	5M	African, European, Hispanic	294	..
NHS	USA	5M	European, Hispanic	314	..
NOMAS	USA	5M	African, European, Hispanic	358	..
REGARDS	USA	5M	African, European, Hispanic	304	..
SPS3	The Americas, Spain	5M	African, European, Hispanic	949	..
SWISS	USA	5M	African, European, Hispanic	181	..
WHI	USA	5M	African, European, Hispanic	454	..
WUSTL	USA	5M	African, European, Hispanic	449	..
HRS	USA	2·5M	African, European, Hispanic	..	11174
OAI	USA	2·5M	African, European	..	3882
HCHS/SOL	USA	2·5M	Hispanic	..	1214
Case-control group 5					
Krakow	Poland	5M	European, Hispanic	880	717
Case-control group 6					
Leuven	Belgium	5M	European, Hispanic	460	453
Case-control group 7					
BASICMAR	Spain	5M	European, Hispanic	890	..
ADHD	Spain	1M	European	..	411
INMA	Spain	1M	European	..	807
Case-control group 8					
GRAZ	Austria	610	European	..	815
GRAZ	Austria	5M	European	609	..
Case-control group 9					
SAHLSIS	Sweden	5M	European, Hispanic	783	..
LUND	Sweden	5M	European, Hispanic	613	..
MDC*	Sweden	610	European, Hispanic	211	1362
Case-control group 10					
ASGC	Australia	610	European	1109	1200
Case-control group 11					
VISP	USA, Canada, UK	1M	African, European	1979	..

(Table 1 continues on next page)

system are improved inter-observer and intra-observer reliability[12,14,15] and the ability to capture and store individual data elements from the clinical evaluation of the subject.

In the first stage of our study, each site assigned stroke subtypes with the CCS system (appendix). All of the CCS data were collected, subjected to quality control, and analysed centrally. Most sites had previously generated TOAST subtype classifications. In the second stage, we identified additional sites that had GWAS data for subtyped stroke cases. Because we included all available CCS-classified cases in the first stage, we used the corresponding subtype categories from TOAST in the second stage.

For both CCS and TOAST, each case was assigned to one of five ischaemic stroke subtypes: cardioembolic, large artery atherosclerosis, small artery occlusion, undetermined, and other. Although the classification of other was available, we did not analyse it because of low sample counts and insufficient power. In CCS, the classification undetermined was used to refer to cryptogenic cases in which no cause was identified after adequate assessment, whereas in TOAST, undetermined cases were those with incomplete assessment, more than one possible cause, and cryptogenic.

Quality control

Full details of the genotyping and quality control processes are provided in the appendix (p 4). Briefly, newly genotyped cases and about 1150 controls were genotyped on the Illumina 5M array (Illumina, San Diego, CA, USA) so we could include them in the analyses for the first stage. All other cases had been previously genotyped on various Illumina platforms (appendix). We selected publicly available external controls to match cases on the basis of ancestral background and genotyping array.

The cases and controls that were newly genotyped formed separate analysis groups (Krakow, Poland, and Leuven, Belgium; table 1). The remaining cases and controls were matched based on cohort, geographic region of the sample collection site, and genotyping platform (table 1). We assigned matched cases and controls into ancestry-specific analysis strata in two steps (appendix). We projected samples onto HapMap 3[16] data using principal component analysis to establish a group of European ancestry samples. Then, we implemented a hyper-ellipsoid clustering technique based on principal components within self-reported groups of non-Hispanic black and Asian participants. We used the hyper-ellipsoid analysis to establish a group of non-Hispanic black (African ancestry) participants and a group of participants of Asian ancestry. Samples that were not grouped as European, African, or Asian ancestry formed the Hispanic stratum. We excluded samples of Asian ancestry from further analysis because of the small number. After establishing the ancestry-based composite groups, we did principal component analysis again to confirm the ancestral homogeneity within each case-control stratum.

Case-control strata then underwent extensive quality control (appendix). Finally, each stratum was prephased[17] and imputed. We imputed samples of European ancestry using a merged reference panel that included the 1000 Genomes Project Phase I[18] and the Genome of the Netherlands.[19] We imputed samples in the African and Hispanic groups using the 1000 Genomes Project Phase I reference panel only. We added summary-level imputed data from an additional cohort (Vitamin Intervention for Stroke Prevention) to the first stage meta-analysis.

First stage genome-wide association analysis

After quality control and imputation, 16 851 cases and 32 473 controls from 15 ancestry-specific groups were available for genome-wide testing (table 1, appendix). Within each stratum, we analysed all ischaemic stroke phenotypes and the four main subtypes (cardioembolism, large artery atherosclerosis, small artery occlusion, and undetermined) as established with CCSc, CCSp, and TOAST, which were available for 12 612 (74·8%) cases. All GWAS were adjusted for sex and the top ten principal components; genome-wide testing was not corrected for age, because age information was missing for most of the controls.

After the GWAS, we removed SNPs with frequency of less than 1% because they showed excessive genomic inflation. We checked the frequencies of imputed SNPs for consistency with the continental populations represented in the 1000 Genomes Project Phase I, and we removed SNPs with a difference in frequency of more than 30%. After quality control, 9·3 million to 15·4 million SNPs were available in the study strata for the meta-analysis. We did inverse variance-weighted fixed-effects meta-analysis across the 15 ancestry-specific strata using MANTEL[20] in each of the 15 traits. The genomic inflation factor λ of the 15 meta-analyses for each trait ranged from 0·936 to 1·005 (appendix pp 5–8).

Second stage analysis

In the second stage, we performed in-silico lookups of association results in 18 independent studies that contained 20 941 TOAST-subtyped cases and 364 736 controls, using the nominally significant SNPs identified in the first stage (table 1 and appendix p 51). The SNPs selected for the second stage for each subtype were aggregated such that, for example, SNPs with $p<1\times10^{-6}$ from the three cardioembolism GWAS (CCSc, CCSp, and TOAST) were all selected for lookup in the independent TOAST cardioembolism cases and matched controls. This process was repeated for the other subtypes.

Joint analysis

We did a meta-analysis of the results from the in-silico lookups from the second stage and the results from the first stage. We set the threshold for genome-wide significance in the joint analysis at $p<1\times10^{-8}$, after correction for testing of the five phenotypes (all stroke,

	Location of sample collection	Genotyping platform	Ancestry groups	Cases	Controls
(Continued from previous page)					
Melanoma Study	USA	1M	European	..	1047
HANDLs	USA	1M	African	..	971
Total	..	..	..	16 851	32 473
Second stage					
ARIC	USA	Affy 6.0	African	263	2466
CADISP†	Multi-cohort	Illumina 610	European	555	9259
CHARGE†	Multi-cohort	Multi-chip	European	3100	75 530
CHS	USA	Illumina Omni 1M	African	110	623
deCODE	Iceland	Multi-chip	European	5291	228 512
Glasgow	UK	ImmunoChip	European	599	1775
HVH	USA	Illumina 370CNV	European	577	1330
INTERSTROKE†	Multi-cohort	Cardio-metabochip	African, East Asian, European, Hispanic	1771	2103
LUND	Sweden	635	European	546	528
MDC	Sweden	5M	European	1304	3504
METASTROKE†	Multi-cohort	Multi-chip	European	1729	7925
RACE	Pakistan	660	South Asian	2385	5193
SAHLSIS	Sweden	750	European	299	596
SIFAP	Germany	2·5M	European	981	1825
SIGNET-REGARDS	USA	Affy 6.0	African	258	2094
SWISS/ISGS	USA	Illumina 610 or 660	African	173	389
UTRECHT	The Netherlands	ImmunoChip	European	556	1145
VHIR-FMT-BARCELONA	Spain	HumanCore and ExomeChip	European	545	320
WGHS‡	USA	Human Hap300 and custom array	European	440	22 725
Total	..	..	..	21 482	367 842
Joint					
Total	..	..	..	38 333	400 315

Case cohorts in the first stage were matched to external controls based on genotyping array, cohort, ancestry, and location of sample collection. Case-control groups were constructed for the first stage analyses from contributing cohorts, which were mainly case-only or control-only cohorts. Hispanic samples were an exception and are not shown as a separate group here, because the small number of samples required that we pool all available Hispanic samples into a single analysis stratum. The second stage consisted of in-silico SNP lookups of summary-level results in previously analysed case-control sets. Totals represent the number of unique samples, accounting for partial sample overlap between two sites (CHARGE and WGHS). NINDS-SiGN=National Institute of Neurological Disorders and Stroke Stroke Genetics Network. BRAINS=Biorepository of DNA in Stroke. MGH-GASROS=Massachusetts General Hospital—Genes Affecting Stroke Risk and Outcome Study. ISGS=Ischemic Stroke Genetics Study. SWISS=Siblings with Ischemic Stroke Study. HABC=Health ABC. EDIN=Edinburgh Stroke Stoke. OXVASC=Oxford Vascular Study. STGEORGE=St George's Hospital. KORA=MONICA/KORA Ausburg Study. WTCCC=Wellcome Trust Case Control Consortium. GEOS=Genetics of Early Onset Stroke. GCNKSS=Greater Cincinnati/Northern Kentucky Stroke Study. MCISS=Middlesex County Ischemic Stroke Study. MIAMISR=Miami Stroke Registry and Biorepository. NHS=Nurses' Health Study. NOMAS=Northern Manhattan Study. REGARDS=Reasons for Geographic and Racial Differences in Stroke. SPS3=Secondary Prevention of Small Subcortical Strokes. WHI=Women's Health Initiative. WUSTL=Washington University St Louis. HRS=Health and Retirement Study. OAI=Osteoarthritis Initiative. HCHS/SOL=The Hispanic Community Health Study/Study of Latinos. LEUVEN=Leuven Stroke Genetics Study. BASICMAR=Base de Datos de Ictus del Hospital del Mar. ADHD=Attention-deficit Hyperactivity Disorder. INMA=Infancia y medio ambiente. SAHLSIS=Sahlgrenska Academy Study of Ischemic Stroke. LUND=Lund Stroke Registry. MDC=Malmo Diet and Cancer Study. ASGC=Australian Stroke Genetics Collaborative. VISP=Vitamin Intervention for Stroke Prevention. HANDLs=Health/Aging in Neighborhoods of Diversity across the Lifespan Study. ARIC=Atherosclerosis Risk in Communities Study. CADISP=Cervical Artery Dissection and Ischemic Stroke Patients. CHARGE=Cohorts for Aging and Research in Genetic Epidemiology. CHS=Cardiovascular Health Study. HVH=Heart and Vascular Health Study. GLASGOW=Glasgow ImmunoChip Study. RACE=Risk Assessment of Cardiovascular Events. SIFAP=Stroke in Young Fabry Patients. SIGNET=The Sea Island Genetics Network. UTRECHT=Utrecht ImmunoChip Study/PROMISe Study. WGHS=Women's Genome Health Study. *Only TOAST subtypes available for the first stage. †Contributing cohorts are described in the appendix. ‡Not included in the ischaemic stroke and cerebroembolism analyses because of overlap with CHARGE.

Table 1: **Case and control cohorts in NINDS-SiGN**

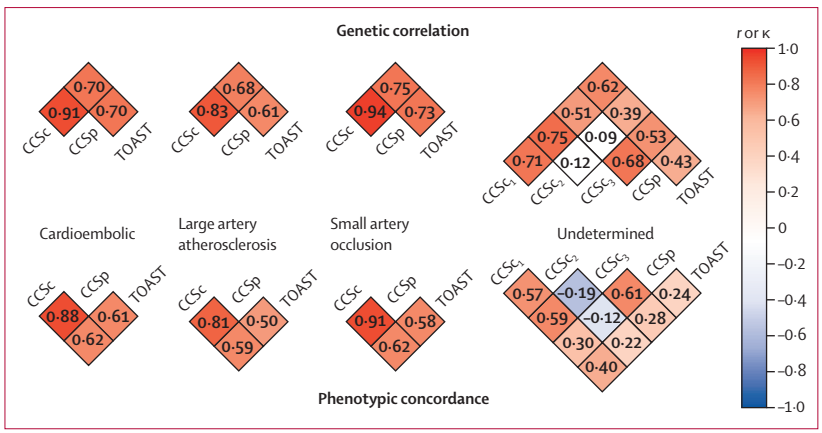


Figure 1: **Genetic and phenotypic correlation between subtyping methods in the first stage analysis**
All cases with an available CCS subtype were included in the first stage analyses. Genome-wide *Z* scores from the CCSc, CCSp, and TOAST GWAS were checked for correlation between each possible pair of traits. Pearson's *r* correlation coefficient (mathematically equivalent in this scenario to the Lin's concordance correlation coefficient) within each square shows genetic correlation. Cohen's κ within each square shows phenotypic agreement. $CCSc_1$ includes all undetermined strokes; $CCSc_2$ includes all incomplete and unclassified strokes; and $CCSc_3$ includes all cryptogenic and cardioembolic minor strokes. The $CCSc_2$ and $CCSc_3$ classifications are mutually exclusive. CCS=Causative Classification of Stroke. CCSc=CCS causative. CCSp=CCS phenotypic. TOAST=Trial of Org 10172 in Acute Stroke Treatment classification system. GWAS=genome-wide association study.

cardioembolic, large artery atherosclerosis, small artery occlusion, and undetermined). λ in the ischaemic stroke joint analysis was 1·005 and ranged from 0·936 to 0·998 in the subtype analyses (appendix pp 9–12).

Role of the funding source

The funder participated in the design of the study. The study investigators were solely responsible for the data collection, analysis, and interpretation. An employee of NINDS (KG) was a member of the writing committee The analysis team had full access to all data included in the study. The steering committee had final responsibility for the decision to submit the report for publication.

Results

After data quality control (appendix p 4 and pp 114–26), we included 16 851 stroke cases and 32 473 controls in the first stage of our analyses. The first stage GWAS revealed 1372 SNPs in 268 loci associated with ischaemic stroke or a specific subtype in any of the CCS or TOAST traits at $p<1\times10^{-6}$. We included an additional independent set of 20 941 cases and 364 736 controls in the second stage, which enabled the joint analysis of 37 893 cases and 397 209 controls across five primary independent traits (ischaemic stroke and the four subtypes).

Genome-wide *Z* scores (SNP β values divided by their respective SE) from the CCSc, CCSp, and TOAST GWAS were checked for correlation (Pearson's *r*) between each possible pair of traits. The analysis revealed moderate to strong genetic correlation (figure 1) between the standardised SNP effects in CCSc, CCSp, and TOAST, despite the modest phenotypic correlation noted previously.[21] The moderate to strong genetic correlation between CCS and TOAST within subtype-specific clusters suggested that TOAST subtyping was appropriate for inclusion in the second stage of the analysis. Phenotypic correlations were also strong within subtype-specific clusters (figure 1).

In the joint analysis of CCS (first stage) and TOAST (second stage) results, SNPs in two novel loci exceeded genome-wide significance. Four common SNPs in linkage disequilibrium ($r^2>0\cdot57$ in the 1000 Genomes Project samples of European ancestry) near the *TSPAN2* locus on chromosome 1 were associated at genome-wide significance with large artery atherosclerosis. The lead SNP in the associated locus was rs12122341 (odds ratio [OR] for the G allele 1·19, 95% CI 1·12–1·26, $p=1\cdot3\times10^{-9}$; figure 2, table 2).

A second locus emerged as having a genome-wide significant association with ischaemic stroke, but only in samples of African ancestry. In view of the small sample size in which it was identified, the association must be interpreted with caution. rs74475935 in *ABCC1* on chromosome 16 was associated with the undetermined phenotype (table 2, appendix p 14), driven by a variant with rare frequency (minor allele frequency [MAF] about 0·01%) in European-ancestry samples and low frequency (MAF about 1·5%) in African-ancestry samples.

We also identified associations for the previously reported loci *PITX2*[4] and *ZFHX3*[5] for cardioembolic stroke, and *HDAC9*[6] for large artery atherosclerotic stroke, all of which exceeded genome-wide significance in our samples (table 2). The 12q24.12 locus near *ALDH2*, previously reported to be associated with all ischaemic stroke, but not with any specific subtype,[7] exceeded genome-wide significance in the joint analysis of all ischaemic stroke (OR for the T allele 1·07, 95% CI 1·5–1·09, $p=4\cdot20\times10^{-9}$). However, the association was even stronger for small artery occlusion in the joint analysis of CCSp in the first stage and TOAST in the second stage (OR 1·17, 95% CI 1·11–1·23, $p=2\cdot92\times10^{-9}$); the association was not genome-wide significant in the joint analysis of CCSc (first stage) and TOAST (second stage; OR 1·16, 95% CI 1·10–1·22, $p=2\cdot77\times10^{-8}$). Evidence of associations with other subtypes was reduced in our study (OR<1·1 and $p>4\times10^{-3}$ for cardioembolism, large artery atherosclerosis, and undetermined in the combined CCSp and TOAST analysis; appendix p 15). Systematic testing that accounted for shared controls (appendix p 15) showed a significant difference in the magnitude of ORs between small artery occlusion and the combined non-small artery occlusion subtypes (p=0·048, appendix p 15), suggesting that the effect of 12q24.12 might be specific for small artery occlusion.

By contrast, we did not find any evidence for the previously reported association between ischaemic stroke and *NINJ2* (rs34166160, OR for the A allele 1·20, 95% CI 0·96–1·48, p=0·106; table 2), even though our sample size had 100% power to detect an association (p<0·05) at this locus. In the full first stage analysis, evidence for association was weak for both the 6p21[22] and

CDKN2B-AS1[23] loci in large artery atherosclerosis, and for the *ABO*[24] locus in all ischaemic stroke, large artery atherosclerosis, and cardioembolism (table 2). When we restricted our analysis to only the samples not used for the initial discovery (appendix p 52), *CDKN2B-AS1* was associated with large artery atherosclerosis (OR for the G allele 1·09, 95% CI 1·02–1·17, p=0·009) and *ABO* was associated with all ischaemic stroke (OR for the C allele 1·07, 95% CI 1·03–1·10, p=$2·5\times10^{-4}$), large artery atherosclerosis (OR 1·15, 95% CI 1·07–1·24, p=$2·5\times10^{-4}$), and cardioembolism (OR 1·09, 95% CI 1·02–1·16, p=0·007). For 6p21, however, we detected no evidence for any association with large artery atherosclerosis (OR for the T allele 1·04, 95% CI 0·96–1·12, p=0·304).

Discussion

Our results show a novel association between a genetic locus and large artery atherosclerosis. The lead SNP, rs12122341, is located in an intergenic region 23·6 kb upstream of *TSPAN2,* the gene encoding tetraspanin-2 (figure 2) This SNP is in linkage disequilibrium with intronic and untranslated region variants in *TSPAN2* (r^2>0·3 in 1000 Genomes Project samples of European ancestry), but is located in a DNA sequence immediately adjacent to *TSPAN2* that can be bound by several transcription factor proteins, including CTCF. This sequence is a promotor and enhancer site that is marked by histone modification and DNase hypersensitivity according to experimental data from ENDCODE and ROADMAP Epigenomics (appendix p 16),[25,26] suggesting a potential role for rs12122341 in gene regulation. An intergenic SNP near rs12122341 has been reported to be associated with migraine,[27] but the two SNPs are not in linkage disequilibrium (r^2=0·03 in 1000 Genomes Project samples of European ancestry).

TSPAN2, the gene closest to rs12122341, is a member of the transmembrane 4 (tetraspanin) superfamily. This family of proteins can mediate signal transduction to regulate cell development, activation, growth, and motility. *TSPAN2* knock-out mice have increased neuroinflammation, shown by activation of microglia and astrocytes with no effect on myelination and axon integrity.[28] Notably, *TSPAN2* is highly expressed in artery tissue and whole blood cells (appendix p 16), which accords with the association we detected between *TSPAN2* with large artery atherosclerosis stroke. Whether the association of rs12122341 is caused by the locus' regulation of *TSPAN2* or other nearby genes will need further functional assessment.

The additional locus that we identified as being associated with undetermined stroke (rs74475935) is in a gene-rich region with linkage-disequilibrium-paired SNPs (r^2>0·1 in 1000 Genomes Project samples of African ancestry) of up to 4 Mb. Because of the small sample size for rs74475935 (610 cases) and the shortage of samples from people with African ancestry, studies with large samples from people of African descent will be necessary to fully assess the robustness of this signal.

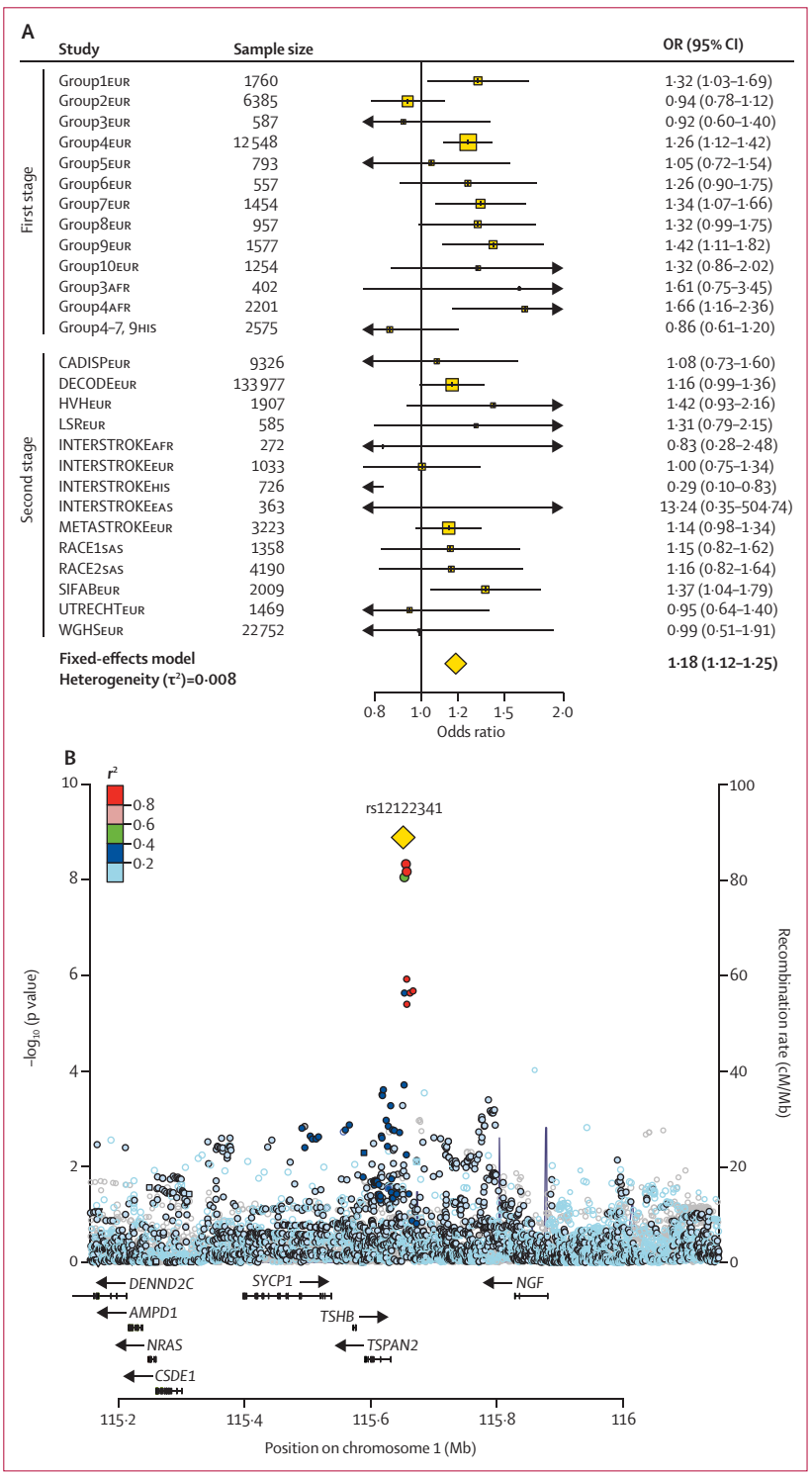


Figure 2: **Forest plot (A) and regional association plot (B) of rs12122341**
Plot of effect size of the association of rs12122341 with large artery atherosclerosis-related stroke across the case–control groups included in the fist and second stage analyses (A). Association of rs12122341 and other SNPs in the region with large artery atherosclerosis-related stroke (B). Point shading shows linkage disequilibrium (r^2) to rs12122341 as calculated in 1000 Genomes Project Phase I European samples. Purple lines show recombination rate. EUR=European ancestry. AFR=African ancestry. HIS=Hispanic samples. EAS=east Asian ancestry. SAS=south Asian ancestry.

	Chromosome	Risk allele	Risk allele frequency (%)			Nearest gene	First stage				Second stage				Joint analysis			
			European	African	The Americas		Subtyping system	Cases	OR (95% CI)	p value	Subtyping system	Cases	OR (95% CI)	p value	Subtyping system	Cases	OR (95% CI)	p value
Novel loci																		
Large artery atherosclerosis																		
rs12122341	1	G	25·7	8·8	19·5	*TSPAN2*	CCSc	2454	1·20 (1·12–1·29)	$3·38×10^{-7}$	TOAST	2249	1·15 (1·04–1·26)	$5·25×10^{-3}$	CCSc	4703	1·18 (1·12–1·25)	$8·32×10^{-9}$
rs12122341	1	G	25·7	8·8	19·5	*TSPAN2*	CCSp	2715	1·21 (1·13–1·30)	$4·50×10^{-8}$	TOAST	2249	1·15 (1·04–1·26)	$5·25×10^{-3}$	CCSp	4964	1·19 (1·12–1·26)	$1·30×10^{-9}$
rs12122341	1	G	25·7	8·8	19·5	*TSPAN2*	TOAST	2346	1·15 (1·07–1·24)	$1·61×10^{-4}$	TOAST	2249	1·15 (1·04–1·26)	$5·25×10^{-3}$	TOAST	4595	1·15 (1·08–1·22)	$2·70×10^{-6}$
Undetermined																		
rs74475935	16	G	0·2	1·8	0·6	*ABCC1*	CCSc	2392*	5·17 (2·99–8·92)	$3·69×10^{-9}$	TOAST	3469	1·87 (0·55–6·41)	$3·16×10^{-1}$	CCSc	5861	4·63 (2·77–7·72)	$4·70×10^{-11}$
rs74475935	16	G	0·2	1·8	0·6	*ABCC1*	CCSp	1062*	8·68 (4·55–16·58)	$5·94×10^{-11}$	TOAST	3469	1·87 (0·55–6·41)	$3·16×10^{-1}$	CCSp	4531	6·89 (3·80–12·47)	$1·85×10^{-10}$
rs74475935	16	G	0·2	1·8	0·6	*ABCC1*	TOAST	3593	2·18 (1·16–4·10)	$1·58×10^{-2}$	TOAST	3469	1·87 (0·55–6·41)	$3·16×10^{-1}$	TOAST	7062	2·11 (1·20–3·70)	$9·22×10^{-3}$
Previously identified loci, first stage $p<1×10^{-6}$																		
All ischaemic stroke																		
rs10744777	12	T	66·7	4·5	5·2	*ALDH2*	..	16 851	1·10 (1·06–1·13)	$3·07×10^{-8}$	..	21 042	1·05 (1·01–1·08)	$6·55×10^{-3}$		37 893	1·07 (1·5–1·09)	$4·20×10^{-9}$
rs2634074	4	T	2·1	4·8	4·1	*PITX2*	..	16 851	1·09 (1·06–1·13)	$2·56×10^{-7}$	..	21 042	1·10 (1·07–1·14)	$2·00×10^{-8}$		37 893	1·10 (1·07–1·12)	$2·68×10^{-14}$
rs2107595	7	A	15·7	2·2	2·2	*HDAC9*	..	16 851	1·10 (1·06–1·14)	$7·74×10^{-7}$	..	21 042	1·07 (1·03–1·11)	$1·70×10^{-4}$		37 893	1·09 (1·06–1·12)	$8·60×10^{-10}$
Cardioembolism																		
rs2200733	4	T	12·0	2·2	2·6	*PITX2*	CCSc	3071	1·39 (1·28–1·50)	$1·24×10^{-16}$	TOAST	3991	1·36 (1·26–1·46)	$1·21×10^{-16}$	CCSc	7062	1·37 (1·30–1·45)	$1·04×10^{-29}$
rs2200733	4	T	12·0	2·2	2·6	*PITX2*	CCSp	3695	1·39 (1·29–1·49)	$3·26×10^{-19}$	TOAST	3991	1·36 (1·26–1·46)	$1·21×10^{-16}$	CCSp	7686	1·37 (1·30–1·45)	$2·79×10^{-32}$
rs2200733	4	T	12·0	2·2	2·6	*PITX2*	TOAST	3427	1·37 (1·27–1·48)	$1·02×10^{-16}$	TOAST	3991	1·36 (1·26–1·46)	$1·21×10^{-16}$	TOAST	7418	1·36 (1·29–1·44)	$8·05×10^{-30}$
rs7193343	16	T	17·4	2·4	18·9	*ZFHX3*	CCSc	3071	1·17 (1·09–1·26)	$1·12×10^{-5}$	TOAST	3991	1·15 (1·07–1·23)	$7·93×10^{-5}$	CCSc	7062	1·17 (1·10–1·22)	$7·28×10^{-9}$

(Table 2 continues on next page)

	Chromosome	Risk allele	Risk allele frequency (%)			Nearest gene	First stage				Second stage				Joint analysis			
			European	African	The Americas		Subtyping system	Cases	OR (95% CI)	p value	Subtyping system	Cases	OR (95% CI)	p value	Subtyping system	Cases	OR (95% CI)	p value
(Continued from previous page)																		
rs7193343	16	T	17·4	2·4	18·9	*ZFHX3*	CCSp	3695	1·19 (1·11–1·27)	$2·93×10^{-7}$	TOAST	3991	1·15 (1·07–1·23)	$7·93×10^{-5}$	CCSp	7686	1·17 (1·11–1·23)	$2·29×10^{-10}$
rs7193343	16	T	17·4	2·4	18·9	*ZFHX3*	TOAST	3427	1·17 (1·09–1·25)	$1·45×10^{-5}$	TOAST	3991	1·15 (1·07–1·23)	$7·93×10^{-5}$	TOAST	7418	1·16 (1·10–1·22)	$8·88×10^{-9}$
Large artery atherosclerosis																		
rs11984041	7	T	9·3	2·2	6·7	*HDAC9*	CCSc	2454	1·30 (1·18–1·42)	$8·46×10^{-8}$	TOAST	2249	1·15 (1·03–1·29)	$1·16×10^{-2}$	CCSc	4703	1·23 (1·15–1·33)	$1·10×10^{-8}$
rs11984041	7	T	9·3	2·2	6·7	*HDAC9*	CCSp	2715	1·29 (1·18–1·42)	$3·50×10^{-8}$	TOAST	2249	1·15 (1·03–1·29)	$1·16×10^{-2}$	CCSp	4964	1·24 (1·15–1·33)	$4·52×10^{-9}$
rs11984041	7	T	9·3	2·2	6·7	*HDAC9*	TOAST	2346	1·30 (1·17–1·43)	$3·62×10^{-7}$	TOAST	2249	1·15 (1·03–1·29)	$1·16×10^{-2}$	TOAST	4595	1·23 (1·14–1·33)	$4·48×10^{-8}$
Small artery occlusion																		
rs10744777	12	T	66·7	4·5	5·2	*ALDH2*	CCSc	2736	1·19 (1·11–1·27)	$9·10×10^{-7}$	TOAST	2426	1·12 (1·03–1·21)	$4·66×10^{-3}$	CCSc	5162	1·16 (1·10–1·22)	$2·77×10^{-8}$
rs10744777	12	T	66·7	4·5	5·2	*ALDH2*	CCSp	2734	1·20 (1·12–1·28)	$6·82×10^{-8}$	TOAST	2426	1·12 (1·03–1·21)	$4·66×10^{-3}$	CCSp	5160	1·17 (1·11–1·23)	$2·92×10^{-9}$
rs10744777	12	T	66·7	4·5	5·2	*ALDH2*	TOAST	3147	1·13 (1·06–1·21)	$1·05×10^{-4}$	TOAST	2426	1·12 (1·03–1·21)	$4·66×10^{-3}$	TOAST	5573	1·13 (1·07–1·18)	$1·62×10^{-6}$
Previously identified loci, first stage $p>1×10^{-6}$																		
All ischaemic stroke																		
rs34166160	12	A	0·9	0·0	0·3	*NINJ2*	..	16 851	1·20 (0·96–1·48)	$1·06×10^{-1}$	..	..	..	..	..	..	..	..
rs11833579	12	G	75·8	79·4	68·0	*NINJ2*	..	16 851	1·02 (0·95–1·01)	$2·15×10^{-1}$	..	..	..	..	..	..	..	..
rs505922	9	C	35·1	32·6	23·5	*ABO*	..	16 851	1·07 (1·04–1·10)	$2·03×10^{-5}$	..	..	..	..	..	..	..	..

(Table 2 continues on next page)

	Chromosome	Risk allele	Risk allele frequency (%)			Nearest gene	First stage				Second stage				Joint analysis			
			European	African	The Americas		Subtyping system	Cases	OR (95% CI)	p value	Subtyping system	Cases	OR (95% CI)	p value	Subtyping system	Cases	OR (95% CI)	p value
(Continued from previous page)																		
Cardioembolism																		
rs505922	9	C	35·1	32·6	23·5	*ABO*	CCSc	3071	1·04 (0·98–1·10)	$1·88×10^{-1}$	..	..	..	..	..	..	..	..
rs505922	9	C	35·1	32·6	23·5	*ABO*	CCSp	3695	1·04 (0·98–1·10)	$1·62×10^{-1}$	..	..	..	..	..	..	..	..
rs505922	9	C	35·1	32·6	23·5	*ABO*	TOAST	3427	1·08 (1·02–1·15)	$5·66×10^{-3}$	..	..	..	..	..	..	..	..
Large artery atherosclerosis																		
rs505922	9	C	35·1	32·6	23·5	*ABO*	CCSc	2454	1·09 (1·02–1·17)	$6·93×10^{-3}$	..	..	..	..	..	..	..	..
rs505922	9	C	35·1	32·6	23·5	*ABO*	CCSp	2715	1·11 (1·04–1·18)	$1·29×10^{-3}$	..	..	..	..	..	..	..	..
rs505922	9	C	35·1	32·6	23·5	*ABO*	TOAST	2346	1·14 (1·06–1·21)	$2·15×10^{-4}$	..	..	..	..	..	..	..	..
rs556621	6	T	29·1	8·1	40·7	6p21	CCSc	2454	1·04 (0·97–1·11)	$3·18×10^{-1}$	..	..	..	..	..	..	..	..
rs556621	6	T	29·1	8·1	40·7	6p21	CCSp	2715	1·02 (0·95–1·19)	$6·36×10^{-1}$	..							
rs556621	6	T	29·1	8·1	40·7	6p21	TOAST	2346	1·11 (1·04–1·19)	$2·55×10^{-3}$	..	..	..	..	..	..	..	..
rs2383207	9	G	49·9	4·5	41·3	*CDKN2B-AS1*	CCSc	2454	1·12 (1·05–1·19)	$4·34×10^{-4}$	..	..	..	..	..	..	..	..
rs2383207	9	G	49·9	4·5	41·3	*CDKN2B-AS1*	CCSp	2715	1·11 (1·05–1·19)	$7·93×10^{-4}$	..	..	..	..	..	..	..	..
rs2383207	9	G	49·9	4·5	41·3	*CDKN2B-AS1*	TOAST	2346	1·09 (1·02–1·17)	$8·13×10^{-3}$	..	..	..	..	..	..	..	..

For subtype-specific loci, ORs and their corresponding p values are reported for the CCSc, CCSp, and TOAST subtypes. Risk allele frequency was calculated with 1000 Genomes (Phase I) European-ancestry samples, African-ancestry samples, and samples from the Americas. Association results were looked up in TOAST-subtyped cases and their matched controls and meta-analysed with the first stage results from CCSc, CCSp, and TOAST cases. CCS=Causative Classification of Stroke. TOAST=Trial of Org 10172 in Acute Stroke Treatment classification system. CCSc=CCS causative. CCSp=CCS phenotypic. OR=odds ratio. *Results from the CCS cryptogenic phenotype.

Table 2: **Novel and previously identified loci implicated in ischaemic stroke and its subtypes through genome-wide testing**

So far, only four loci—*PITX2*,[4] *ZFHX3*,[5] *HDAC9*,[6] and 12q24.12[7]—have been repeatedly identified in GWAS of ischaemic stroke, all of which are subtype specific except for 12q24.12. Although the 12q locus association was originally identified for all ischaemic stroke, our analysis suggests that it is probably specific to small artery occlusion. These findings suggest that ischaemic stroke subtypes have distinct genetic signatures. Our analysis of genetic correlation across the traits also showed that the subtypes share subtle genetic associations (appendix p 17 and p 53). This finding is supported by the results of another study, which identified genetic overlap between the large artery atherosclerosis and small artery occlusion subtypes.[29] Future efforts will help to clarify both the shared and unique genetic architectures within and between subtypes.

Until now, GWAS of ischaemic stroke subtypes have used far smaller sample sizes than studies of other complex traits. The SiGN study, the largest GWAS of ischaemic stroke so far, was well powered (75·1%) to detect common SNP subtype-specific associations of larger effect (MAF 25% and OR 1·2 in 3000 cases and 30 000 controls) but was substantially less powered to identify lower frequency or lower effect SNPs (13·8% power for MAF 10% and OR 1·2; 1·1% power for MAF 25% and OR 1·1). Because of the almost linear relation that exists between sample size and discovered loci,[30] and because large-scale GWAS in other complex traits have yielded hundreds of SNP-disease associations,[31–33] studying ischaemic stroke subtypes in larger samples will probably yield additional associated common variants. Furthermore, the implementation of whole genome sequencing studies of stroke will begin to test whether rare alleles in the population account for a substantial proportion of disease heritability.

The SiGN study has several other limitations. First, sample inclusion was heavily biased towards individuals of European descent; inclusion of non-European populations will improve power for locus discovery[34] and will be especially informative for future fine-mapping efforts.[35] Second, the inclusion of TOAST-based classification for samples in the second stage probably added phenotypic heterogeneity (figure 1, appendix p 53), which potentially reduced power.[21] Third, many of the participating studies within SiGN (especially the publicly available controls) had little or no stroke-specific risk factor data available. Such data are key to disentangling potential gene–environment interactions. Future genetic studies of stroke will continue to face challenges related to the disease phenotype, including high prevalence of the disease (lifetime risk about 20%), its late onset (mainly in individuals >65 years), the contribution of other cardiovascular diseases and environment as causative factors, and difficulties of subtyping (in SiGN 12·6–22·3% of all cases analysed were ultimately classified as undetermined by CCS or TOAST).

Our use of CCS enabled identification of candidate SNPs that were not significant for the second stage follow-up in TOAST, including those SNPs at the *TSPAN2* locus. This refinement might represent a reduction in phenotypic heterogeneity that CCS introduces through its capture of clinical stroke features, completeness of diagnostic investigations, and, where possible, classification of cases with different potential causes into the most probable causes. The association signal of the *TSPAN2* locus identified with CCS was, however, improved by the inclusion of TOAST-classified samples, suggesting that making use of the genetic correlation underlying the subtyping methods and allowing for broader inclusion of cases, regardless of subtyping system, can lead to the discovery of more susceptibility loci. Further studies will help to establish whether the rich repository of individual-level data created through the use of the CCS will help to uncover novel phenotypes and thus reveal biological mechanisms and broaden the understanding of the genetic architecture in patients with stroke.

NINDS Stroke Genetics Network (SiGN) and International Stroke Genetics Consortium (ISGC)

Sara L Pulit*, Patrick F McArdle*, Quenna Wong*, Rainer Malik*, Katrina Gwinn, Sefanja Achterberg, Ale Algra, Philippe Amouyel, Christopher D Anderson, Donna K Arnett, Ethem Murat Arsava, John Attia, Hakan Ay, Traci M Bartz, Thomas Battey, Oscar R Benavente, Steve Bevan, Alessandro Biffi, Joshua C Bis, Susan H Blanton, Giorgio B Boncoraglio, Robert D Brown Jr, Annette I Burgess, Caty Carrera, Sherita N Chapman Smith, Daniel I Chasman, Ganesh Chauhan, Wei-Min Chen, Yu-Ching Cheng, Michael Chong, Lisa K Cloonan, John W Cole, Ioana Cotlarciuc, Carlos Cruchaga, Elisa Cuadrado-Godia, Tushar Dave, Jesse Dawson, Stéphanie Debette, Hossein Delavaran, Cameron A Dell, Martin Dichgans, Kimberly F Doheny, Chuanhui Dong, David J Duggan, Gunnar Engström, Michele K Evans, Xavier Estivill Pallejà, Jessica D Faul, Israel Fernández-Cadenas, Myriam Fornage, Philippe M Frossard, Karen Furie, Dale M Gamble, Christian Gieger, Anne-Katrin Giese, Eva Giralt-Steinhauer, Hector M González, An Goris, Solveig Gretarsdottir, Raji P Grewal, Ulrike Grittner, Stefan Gustafsson, Buhm Han, Graeme J Hankey, Laura Heitsch, Peter Higgins, Marc C Hochberg, Elizabeth Holliday, Jemma C Hopewell, Richard B Horenstein, George Howard, M Arfan Ikram, Andreea Ilinca, Erik Ingelsson, Marguerite R Irvin, Rebecca D Jackson, Christina Jern, Jordi Jiménez Conde, Julie A Johnson, Katarina Jood, Muhammad S Kahn, Robert Kaplan, L Jaap Kappelle, Sharon L R Kardia, Keith L Keene, Brett M Kissela, Dawn O Kleindorfer, Simon Koblar, Daniel Labovitz, Lenore J Launer, Cathy C Laurie, Cecelia A Laurie, Cue Hyunkyu Lee, Jin-Moo Lee, Manuel Lehm, Robin Lemmens, Christopher Levi, Didier Leys, Arne Lindgren, W T Longstreth Jr, Jane Maguire, Ani Manichaikul, Hugh S Markus, Leslie A McClure, Caitrin W McDonough, Christa Meisinger, Olle Melander, James F Meschia, Marina Mola-Caminal, Joan Montaner, Thomas H Mosley, Martina Müller-Nurasyid, Mike A Nalls, Jeffrey R O'Connell, Martin O'Donnell, Ángel Ois, George J Papanicolaou, Guillaume Paré, Leema Reddy Peddareddygari, Annie Pedersén, Joanna Pera, Annette Peters, Deborah Poole, Bruce M Psaty, Raquel Rabionet, Miriam R Raffeld, Kristiina Rannikmäe, Asif Rasheed, Petra Redfors, Alex P Reiner, Kathryn Rexrode, Marta Ribasés, Stephen S Rich, Wim Robberecht, Ana Rodriguez-Campello, Arndt Rolfs, Jaume Roquer, Lynda M Rose, Daniel Rosenbaum, Natalia S Rost, Peter M Rothwell, Tatjana Rundek, Kathleen A Ryan, Ralph L Sacco, Michèle M Sale, Danish Saleheen, Veikko Salomaa, Cristina Sánchez-Mora, Carsten Oliver Schmidt, Helena Schmidt, Reinhold Schmidt, Markus Schürks, Rodney Scott, Helen C Segal, Stephan Seiler, Sudha Seshadri, Pankaj Sharma, Alan R Shuldiner, Brian Silver, Agnieszka Slowik, Jennifer A Smith, Martin Söderholm, Carolina Soriano, Mary J Sparks, Tara Stanne, Kari Stefansson, O Colin Stine, Konstantin Strauch, Jonathan Sturm, Cathie LM Sudlow, Salman M Tajuddin, Robert L Talbert, Turgut Tatlisumak, Vincent Thijs, Gudmar Thorleifsson, Unnur Thorsteindottir, Steffen Tiedt, Matthew Traylor, Stella Trompet, Valerie Valant, Melanie Waldenberger,

Matthew Walters, Liyong Wang, Sylvia Wassertheil-Smoller, David R Weir, Kerri L Wiggins, Stephen R Williams, Dorota Wloch-Kopec, Daniel Woo, Rebecca Woodfield, Ona Wu, Huichun Xu, Alan B Zonderman, Australian Stroke Genetics Consortium, Cervical Artery Dissection and Ischemic Stroke Patients (CADISP) study, Cohorts of Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, Consortium of Minority Population genome-wide Association Studies of Stroke (COMPASS), METASTROKE consortium, Wellcome Trust Case-Control Consortium, Bradford B Worrall*, Paul IW de Bakker*, Steven J Kittner*, Braxton D Mitchell*, Jonathan Rosand*. *Contributed equally.
NINDS Stroke Genetics Network (SiGN) and International Stroke Genetics Consortium Writing committee Jonathan Rosand (chair), Braxton D Mitchell (co-chair), Hakan Ay, Paul I W de Bakker, Katrina Gwinn, Steven J Kittner, Arne Lindgren, James F Meschia, Sara L Pulit, Cathie L M Sudlow, Vincent Thijs, Daniel Woo, Bradford B Worrall. *Steering committee* Donna K Arnett, Oscar Benavente, John W Cole, Martin Dichgans, Raji P Grewal, Christina Jern, Jordi Jiménez Conde, Julie A Johnson, Steven J Kittner, Jin-Moo Lee, Christopher Levi, Arne Lindgren, Hugh S Markus, Olle Melander, James F Meschia, Kathryn Rexrode, Jonathan Rosand, Peter M Rothwell, Tatjana Rundek, Ralph L Sacco, Reinhold Schmidt, Pankaj Sharma, Agnieszka Slowik, Cathie L M Sudlow, Vincent Thijs, Sylvia Wasssertheil-Smoller, Daniel Woo, Bradford B Worrall.

Contributors

JR, BDM, HA, PIWdB, SJK, AL, JFM, SLP, CLMS, VT, DW, and BBW contributed to data collection and provided critical review of the manuscript. JR, BDM, HA, PIWdB, KG, SJK, AL, SLP, CLMS, VT, DW, and BBW made critical decisions regarding study design and conduct. JR, BDM, HA, PIWdB, KG, SJK, AL, SLP, CLMS, VT, DW, and BBW participated in literature search and writing of the paper. BDM, PIWdB, and SLP did the statistical analysis and data interpretation.

Declaration of interests

KG is an employee of NINDS. The other members of the writing committee declare no competing interests.

Acknowledgments

The SiGN study was funded by a cooperative agreement grant from the US National Institute of Neurological Disorders and Stroke, National Institutes of Health (U01 NS069208). The information about funding for each collection is reported in the appendix (pp 159–70).

References

1 World Health Organization. The top 10 causes of death. May 2014. http://www.who.int/mediacentre/factsheets/fs310/en/ (accessed April 23, 2015).
2 Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation* 2014; **131:** e29–322.
3 Meschia JF, Bushnell C, Boden-Albala B, et al. Guidelines for the primary prevention of stroke: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2014; **45:** 3754–832.
4 Gretarsdottir S, Thorleifsson G, Manolescu A, et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol* 2008; **64:** 402–09.
5 Gudbjartsson DF, Holm H, Gretarsdottir S, et al. A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet* 2009; **41:** 876–78.
6 Bellenguez C, Bevan S, Gschwendtner A, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet* 2012; **44:** 328–33.
7 Kilarski LL, Achterberg S, Devan WJ, et al. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology* 2014; **83:** 678–85.
8 Bevan S, Traylor M, Adib-Samii P, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke* 2012; **43:** 3161–67.
9 Meschia JF, Arnett DK, Ay H, et al. Stroke Genetics Network (SiGN) study: design and rationale for a genome-wide association study of ischemic stroke subtypes. *Stroke* 2013; **44:** 2694–702.
10 Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; **38:** 209–13.
11 Adams H, Bendixen B, Kappelle L, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 1993; **24:** 35–41.
12 Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: the causative classification of stroke system. *Stroke* 2007; **38:** 2979–84.
13 Kolominsky-Rabas PL, Weber M, Gefeller O, Neundoerfer B, Heuschmann PU. Epidemiology of ischemic stroke subtypes according to TOAST criteria: incidence, recurrence, and long-term survival in ischemic stroke subtypes: a population-based study. *Stroke* 2001; **32:** 2735–40.
14 Arsava EM, Ballabio E, Benner T, et al. The causative classification of stroke system: an international reliability and optimization study. *Neurology* 2010; **75:** 1277–84.
15 Ay H, Arsava EM, Andsberg G, et al. Pathogenic ischemic stroke phenotypes in the NINDS-Stroke Genetics Network. *Stroke* 2014; **45:** 3589–96.
16 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38:** 904–09.
17 Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 2012; **44:** 955–59.
18 Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491:** 56–65.
19 Francioli LC, Menelaou A, Pulit SL, et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 2014; **46:** 818–25.
20 de Bakker PIW, Ferreira MAR, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* 2008; **17:** R122–28.
21 McArdle PF, Kittner SJ, Ay H, et al. Agreement between TOAST and CCS ischemic stroke classification: the NINDS SiGN study. *Neurology* 2014; **83:** 1653–60.
22 Holliday EG, Maguire JM, Evans T-J, et al. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet* 2012; **44:** 1147–51.
23 Smith JG, Melander O, Lövkvist H, et al. Common genetic variants on chromosome 9p21 confers risk of ischemic stroke a large-scale genetic association study. *Circ Cardiovasc Genet* 2009; **2:** 159–64.
24 Williams FMK, Carter AM, Hysi PG, et al. Ischemic stroke is associated with the ABO locus: the EuroCLOT study. *Ann Neurol* 2013; **73:** 16–31.
25 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015; **518:** 317–30.
26 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; **40:** D930–34.
27 Esserlind AL, Christensen AF, Le H, et al. Replication and meta-analysis of common variants identifies a genome-wide significant locus in migraine. *Eur J Neurol* 2013; **20:** 765–72.
28 de Monasterio-Schrader P, Patzig J, Möbius W, et al. Uncoupling of neuroinflammation from axonal degeneration in mice lacking the myelin protein tetraspanin-2. *Glia* 2013; **61:** 1832–47.
29 Holliday EG, Traylor M, Malik R, et al. Genetic overlap between diagnostic subtypes of ischemic stroke. *Stroke* 2015; **46:** 615–19.
30 Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012; **90:** 7–24.
31 Ripke S, Neale BM, Corvin A, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511:** 421–27.
32 Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014; **46:** 1173–86.
33 Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013; **45:** 1274–83.
34 Pulit SL, Voight BF, de Bakker PIW. Multiethnic genetic association studies improve power for locus discovery. *PLoS One* 2010; **5:** e12600.
35 Zaitlen N, Paşaniuc B, Gur T, Ziv E, Halperin E. Leveraging genetic variability across populations for the identification of causal variants. *Am J Hum Genet* 2010; **86:** 23–33.

5. Discussion

5.1. Comprehensive summary of the results

This work integrates findings from two unbiased screening approaches, genomics and miRNA profiling, as well as from the largest stroke etiology classification analysis to date to estimate the value of different diagnostic tools and to identify novel prospects for diagnostic approaches in patients with IS.

First, we used RNA sequencing and qRT-PCR in three independent samples to study expression changes of circulating miRNAs after IS. Three out of 32 initially identified miRNAs were validated and replicated: miR-125a-5p, miR-125b-5p and miR-143-3p. Their expression levels depended on platelet numbers in spike-in experiments, but were neither associated with infarct volume or stroke etiology nor affected after neuronal injury *in vitro* or in two different experimental stroke models. When used to discriminate between patients with acute IS and HCs this combination of miRNAs showed an unprecedented diagnostic utility for a blood-based biomarker and outperformed multimodal CT in our sample.

Next, we classified the etiology of 16.267 patients with IS from 20 European and US sites according to two diagnostic classification systems, TOAST and CCS, both integrating clinical findings and technical studies. The agreement was moderate ($\kappa = 0.59$). Both systems could not determine stroke etiology in approximately 25 % of patients.

Finally, investigating the association of genetic variants with IS in 37.893 patients with IS and 397.209 controls, we identified a novel locus near *TSPAN2* being associated with large artery atherosclerosis-related stroke. Four previously identified loci, *HDAC9* for large artery atherosclerosis-related stroke, *PITX2* and *ZFHX3* for cardioembolic stroke, and 12q24 near *ALDH2* for small-artery stroke were confirmed.

Integrating the results of these three studies highlights the potential of innovative blood-based biomarkers for the diagnosis of IS as well as the need to explore their value in other diagnostic questions posed in patients with IS, which are currently not sufficiently answered by clinical findings and technical studies, e.g. by following up novel GWAS signals.

5.2. Translation of the current findings into clinical routine

The number of studies investigating blood-based biomarkers for IS has grown 10-fold over the last 20 years (Figure 1), but so far none of these has led to guideline-supported implementation of a blood-based biomarker test into clinical routine.

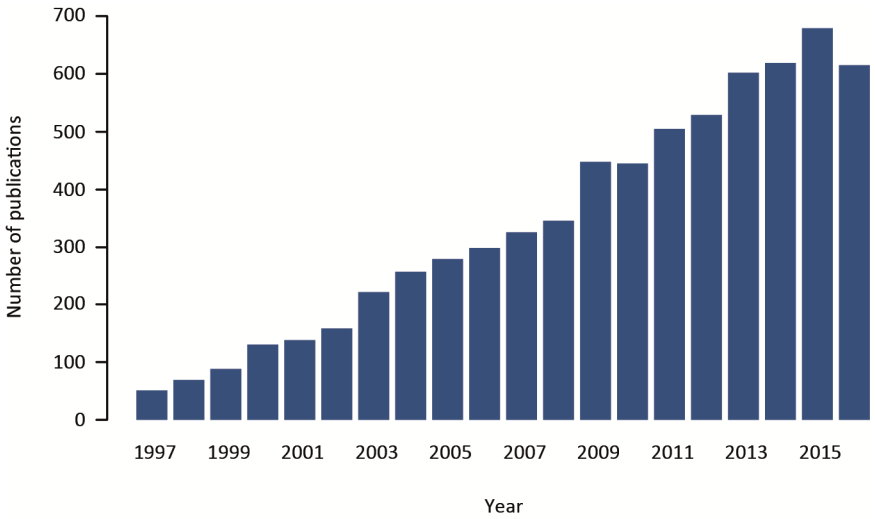


Figure 1. Number of publications investigating blood-based biomarker for IS
A literature search in PubMed was queried on September 28, 2017, using the search terms "blood AND biomarker AND stroke".

A number of prerequisites have been identified for the implementation of a blood-based biomarker test into clinical routine. The concept of *unmet clinical needs* illustrates the primary need to identify a missing or inadequately performing component in a standard process of care.[53] It includes the recognition of the target group, of the limitations of current practice, and of desired outcomes as well as exclusion of alternatives such as optimization of current practice.[53] Identification of desirable clinical performance criteria such as clinical validity and diagnostic accuracy will guide the design of the study including appropriate validation.[54] The selection process for a biomarker (set) should then integrate understanding of the biological processes leading to higher clinical performance and effectiveness of the selected biomarker. Clinical implementation also requires sufficient analytical performance of the biomarker test system including but not limited to limit of detection and quantitation, measurement range, and linearity.[54] Depending on the respective unmet clinical need as well as the selected clinical performance criteria, the economical and technical feasibility of the test will further guide the process of its clinical implementation.

In *RNA-Seq Identifies Circulating miR-125a-5p, miR-125b-5p and miR-143-3p as Potential Biomarkers for Acute Ischemic Stroke* we primarily aimed at identifying biomarkers to diagnose patients with IS at hospital arrival. CT is the current gold standard for diagnostic evaluation in the acute phase of suspected IS, but does not show signs of ischemia in up to 50 % of cases.[55] Still, patients would benefit from accurate selection for further management on the stroke unit, rehabilitative measures and secondary prevention. Using a case-control design in a three-step approach with independent validation and replication samples, we identified three platelet-derived miRNAs, which have also been linked to phenotype and function of vascular smooth muscle cells[56] and endothelial cells[57] as well as to angiogenesis.[58,59] The clinical performance of the set of these three miRNAs was unprecedented for a blood-based biomarker when differentiating patients with IS from HCs. The diagnostic utility for the differentiation of patients with IS and patients with TIA, an important differential diagnosis to IS at hospital arrival, was lower and may not be sufficient for answering this question in clinical routine. In this study, we used qRT-PCR, the current gold standard for quantifying miRNA expression levels. However, using different isolation techniques,[60] primer designs,[61] and qPCR machines may lead to different results between laboratories. Due to the confounding effects of medication[62] and comorbidities[63] our analysis included multivariate adjustment, which is difficult to accomplish in a clinical setting. Also, to date, the process of RNA isolation from platelet-poor plasma samples, cDNA synthesis and qPCR takes a minimum of five hours and is almost always performed on a group- and not individual-based level. Emerging other techniques include the single molecule array platform SIMOA, employing single detection of magnetic beads coupled to miRNA detection probes,[64] TORNADO, a technology based on electrocatalytic platinum nanoparticles,[65] a microchip-based double-strand ligation assay,[66] and base-stacking isothermal amplification.[67] All four approaches quantify miRNAs levels directly, thereby saving time by skipping RNA isolation and cDNA synthesis, but are still in the developmental phase and have not been shown to be applicable in a clinical context. Hence, the identified miRNAs are biologically linked to IS and show a promising clinical performance for an important clinical question with an existing gap in diagnostics. However, many confounders and high variability in the gold standard technique preclude clinical implementation in the near future. Prospective multicenter trials as well as technical advancements will guide the translation of replicated biomarkers.[68]

Correct etiological classification of patients with IS leads to more accurate secondary prophylactic treatment and counsel. The widely used TOAST classification system categorizes approximately 30 % of patients with IS as "undetermined".[15] Initial data on CCS not only suggested its use as a web-based research tool, but also showed a promising reduction of patients categorized as undetermined etiology in a small sample.[16,69,70] However, in *Agreement between TOAST and CCS ischemic stroke classification: the NINDS SiGN Study* we demonstrated that also the CCS rates patients with IS in approximtely 25 % as undetermined. It becomes increasingly evident that the integration of the currently available most specific and sensitive clinical findings and technical studies is not sufficient to provide accurate etiological classification eventually guiding secondary prophylactic treatment.

To date, genetic testing is only performed in IS patients with suspected rare mutations causing Mendelian diseases like Fabry's disease. Being eligible for enzyme replacement therapy,[71] the identification of these patients by genetic testing leads to specific treatment and improves their longterm outcome. In contrast, it remains unclear, which consequences should be drawn from the knowledge about an individual's genetic polymorphisms. Whether assessing the genotype could assist in therapeutic decisions has probably best been investigated for the question whether genotype-guided compared to clinically guided dosing of warfarin improves patients' safety outcomes. Whereas initial studies investigating the outcome parameter time within therapeutic range[72,73] were controversial, a recent multicenter randomized trial revealed results in favor of genotype-guided dosing.[74] The identification of IS subgroup-specific loci has not led to the implementation of genetic testing in a broader population of patients with IS yet. Although sequencing costs have dropped significantly in recent years, the effect sizes for subtype-associations appear too small, to base risk stratification or etiological classification solely on polymorphisms. However, the identification of these loci has also prompted experimental studies investigating their functional link to stroke[75] and may eventually lead to new therapeutic interventions.

5.3. Future directions of diagnostic testing in patients with IS

All four questions raised in section 3.1.3. have identified existing gaps of current diagnostic tests in stroke care. In the following paragraphs I will present a selection of biomarkers to

address these questions combined with necessary question-specific performance parameters, e.g. sensitivity and specificity, but also duration to result.

Shortening the time between symptom onset and initiation of thrombolysis in patients with IS is associated with better outcome.[76] Currently, neuroimaging to rule out intracranial bleeding is the only technical study that is required in patients otherwise eligible for rt-PA administration. Although a CT-equipped mobile stroke treatment unit has been implemented in some metropolitan areas,[77] its application in rural areas is unlikely.[78] Optimization of in-hospital stroke care including hospital prenotification, premixing of rt-PA and reduced door-to-imaging, door-to-needle time can be reduced to a median of 20 minutes.[79] Enabling thrombolysis prior to hospital arrival by means of a blood-based biomarker test that differentiates between IS and HS in the pre-hospital setting would reduce the symptom-to-needle time even further. Applying point-of-care testing would lead to immediate output of results and simple use. Given the risks of applying thrombolytic treatment in patients with HS, this test needs to exclude all patients with HS (100 % sensitivity for HS, specificity for IS, positive-predictive value for IS, or negative-predictive value for HS) and identify a high proportion of patients with IS (high sensitivity for IS). Most, if not all studies have focused on how to exclude patients with HS and investigated serum levels of GFAP, a marker of astrocytic injury, increasing very early after HS.[21] However, shortly after symptom onset, when the benefit from thrombolysis is largest, or in patients with smaller hemorrhage, serum levels of GFAP are low. Accordingly, applying a cut-off to rule out all patients with HS (100 % sensitivity for HS) leads to very low specificity for HS and sensitivity for IS.[23] More favorable cut-offs may be achieved using ultra-sensitive assays like SIMOA or by combining GFAP with other astrocytic markers like the glutamate aspartate transporter, glutamine synthetase, aquaporin 4 or ALDH1L1.[80] Also, combining the HS marker GFAP with a marker for IS could yield higher sensitivity for IS along with 100 % specificity for IS. Early biomarkers of IS could include markers of platelet aggregation, thrombus formation, or hypoxia. With improving door-to-needle and symptom-onset-to-needle times in first-world countries over time and seemingly narrowing windows for the application of such a biomarker, it is critical not to forget that this biomarker test would have enormous health care and economical value for countries with less developed stroke care.

In patients with occlusions of large cerebral arteries, MT has over a short period of time become a mainstay in these patients' care.[7] Outcome is influenced by numerous parameters including time from symptom onset to arrival of the emergency team, collateral status, distance to the next local stroke unit, and distance to the next specialized stroke center with available endovascular service. The decision whether the patient should be transferred to a local stroke unit to enable immediate thrombolytic therapy or to a specialized stroke center for faster potential MT and accepting delayed thrombolytic therapy depends on these parameters but also on the early identification of these patients.[81,82] However, currently, there is no predictor of occlusions of arteries of the proximal cerebral circulation in the preclinical setting. In this context, blood-based biomarkers could be of value as surrogate markers for thrombus size, indicating large vessel occlusions, and for the extent of lost and salvageable brain tissue, resulting from the combination of time from symptom onset to arrival of the emergency team, collateral status and the vessel occlusion itself.

The extent of lost brain tissue has been shown to highly correlate with "destruction markers" such as levels of neuron-specific enolase and S100 calcium binding protein B in the subacute phase but not in the acute phase.[83] However, it has not been investigated whether these markers of cerebral damage could be of value in the acute phase when measured with up-to-date ultrasensitive assays. Also, a blood biomarker as a surrogate marker for the tissue at risk, the penumbra, has not been identified. Few studies investigated markers for oxidative stress, but have not correlated these results to penumbral size.[84] Various potential circulating surrogate markers for platelet activation have been identified, among them the soluble CD40 ligand,[85] activated glycoprotein IIb/IIIa, P-selectin and platelet microparticles,[86] but none of them has been correlated with either the presence of a large vessel occlusion, thrombus size, or the size of the occluded vessel. Notably, we here identified three platelet-associated miRNAs being elevated in the acute phase of IS. Future analyses will show whether these or other platelet-enriched miRNAs, e.g. miR-223-3p or miR-126-3p,[62,87] are associated with large vessel occlusion in patients with IS.

IS is one of many causes of acute neurological dysfunction. In addition to HS, these also include so-called stroke mimics and include migraine, epileptic seizures, and infections among others. However, only few of them are ruled out by neuroimaging before starting thrombolysis. Given the lack of accurate and timely diagnostic tests to rule out stroke mimics,

current clinical routine instructs physicians to risk thrombolysing also an unknown number of patients with stroke mimics for the sake of the benefit of faster thrombolysis in patients with IS.[88] Given this current practice, the biomarker test needs to show 100 % sensitivity for IS (matching 100 % specificity for stroke mimics or 100 % positive predictive value for stroke mimics). Given the heterogeneity of diseases mimicking stroke, the selection of a unifying biomarker for stroke mimics may be more challenging compared to the definition of a lower threshold of a marker for IS. Accordingly, patients missing this lower threshold would be excluded from thrombolytic treatment. The biomarker could involve circulating endothelial cells[89,90] or any of the already specified markers for cerebral damage, oxidative stress, or platelet activation.

Although IS can be caused by many different etiologies categorized within classification systems like TOAST and CCS, the clinically most relevant question with therapeutic consequences to date is to identify patients with AF as they benefit from oral anticoagulation. In contrast, patients with IS categorized as large vessel stroke, small-artery occlusion, and undetermined etiology benefit from secondary prevention with the antiplatelet agent acetylsalicylic acid.[91] Ongoing multicentric randomized trials, RESPECT-ESUS, NAVIGATE-ESUS, and ATTICUS now investigate whether the anticoagulating agents Dabigatran, Rivaroxaban, or Apixaban, respectively, in comparison to the antiplatelet agent acetylsalicylic acid show longer time to first recurrent stroke also in patients with recent embolic stroke of unknown source (ESUS). ESUS is defined as a non-lacunar brain infarct without proximal arterial stenosis or cardioembolic sources[92] and represents a subgroup of patients formerly classified as undetermined. The results of these trials will enhance our understanding about the potential source of thromboembolism in the individual patient based upon which patient subgroup benefits from anticoagulation and which not. The integration of these findings with new pathophysiological concepts like the thrombogenic atrial substrate[93] may lead to new approaches using blood-based biomarkers to define etiological subgroups of IS.

Aside from these short- to medium-term perspectives, long-term perspectives about the future of blood-based diagnostic testing include the application of (multi)-omics techniques like the investigation of DNA, RNA, proteins, and metabolites as point-of-care systems in the preclinical setting as well as at the bedside with visualization tools for the physician beyond

lists and tables of numbers. The achievement to monitor heart rate and blood pressure among other parameters continuously *in vivo* with arterial lines, illustrates that also continuous blood monitoring could not be too far-fetched. Eventually, this would also allow for faster disease detection and maybe even to identify patients at risk for IS or secondary deterioration after IS.

6. References

1. Sacco, R.L.*, et al.* An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke; a journal of cerebral circulation* **44**, 2064-2089 (2013).
2. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1459-1544 (2016).
3. Meschia, J.F.*, et al.* Guidelines for the primary prevention of stroke: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke; a journal of cerebral circulation* **45**, 3754-3832 (2014).
4. Stoll, G., Kleinschnitz, C. & Nieswandt, B. Molecular mechanisms of thrombus formation in ischemic stroke: novel insights and targets for treatment. *Blood* **112**, 3555-3562 (2008).
5. Moskowitz, M.A., Lo, E.H. & Iadecola, C. The science of stroke: mechanisms in search of treatments. *Neuron* **67**, 181-198 (2010).
6. Dirnagl, U., Iadecola, C. & Moskowitz, M.A. Pathobiology of ischaemic stroke: an integrated view. *Trends Neurosci* **22**, 391-397 (1999).
7. Goyal, M.*, et al.* Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet* **387**, 1723-1731 (2016).
8. Flossmann, E., Schulz, U.G. & Rothwell, P.M. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke; a journal of cerebral circulation* **35**, 212-227 (2004).
9. Boehme, A.K., Esenwa, C. & Elkind, M.S. Stroke Risk Factors, Genetics, and Prevention. *Circulation research* **120**, 472-495 (2017).
10. Gretarsdottir, S.*, et al.* Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Annals of neurology* **64**, 402-409 (2008).
11. Gudbjartsson, D.F.*, et al.* A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nature genetics* **41**, 876-878 (2009).
12. Bellenguez, C.*, et al.* Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nature genetics* **44**, 328-333 (2012).
13. Kilarski, L.L.*, et al.* Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology* **83**, 678-685 (2014).
14. Adams, H.P., Jr.*, et al.* Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke; a journal of cerebral circulation* **24**, 35-41 (1993).
15. Grau, A.J.*, et al.* Risk factors, outcome, and treatment in subtypes of ischemic stroke: the German stroke data bank. *Stroke; a journal of cerebral circulation* **32**, 2559-2566 (2001).

16. Ay, H.*, et al.* An evidence-based causative classification system for acute ischemic stroke. *Annals of neurology* **58**, 688-697 (2005).

17. Hennerici, M.G., Kern, R. & Neurologie, K.L.d.D.G.f. S1-Leitlinie Diagnostik akuter zerebrovaskulärer Erkrankungen. *Deutsche Gesellschaft für Neurologie, Hrsg. Leitlinien für Diagnostik und Therapie in der Neurologie.* (2017).

18. Muir, K.W., Buchan, A., von Kummer, R., Rother, J. & Baron, J.C. Imaging of acute stroke. *Lancet neurology* **5**, 755-768 (2006).

19. Sanna, T.*, et al.* Cryptogenic stroke and underlying atrial fibrillation. *N Engl J Med* **370**, 2478-2486 (2014).

20. Brambatti, M.*, et al.* Temporal relationship between subclinical atrial fibrillation and embolic events. *Circulation* **129**, 2094-2099 (2014).

21. Brunkhorst, R., Pfeilschifter, W. & Foerch, C. Astroglial proteins as diagnostic markers of acute intracerebral hemorrhage-pathophysiological background and clinical findings. *Transl Stroke Res* **1**, 246-251 (2010).

22. Ren, C.*, et al.* Assessment of Serum UCH-L1 and GFAP in Acute Stroke Patients. *Scientific reports* **6**, 24588 (2016).

23. Rozanski, M.*, et al.* Glial Fibrillary Acidic Protein for Prehospital Diagnosis of Intracerebral Hemorrhage. *Cerebrovascular diseases* **43**, 76-81 (2017).

24. Foerch, C.*, et al.* Diagnostic accuracy of plasma glial fibrillary acidic protein for differentiating intracerebral hemorrhage and cerebral ischemia in patients with symptoms of acute stroke. *Clinical chemistry* **58**, 237-245 (2012).

25. Katsanos, A.H.*, et al.* Plasma Glial Fibrillary Acidic Protein in the Differential Diagnosis of Intracerebral Hemorrhage. *Stroke; a journal of cerebral circulation* **48**, 2586-2588 (2017).

26. Anand, N. & Stead, L.G. Neuron-specific enolase as a marker for acute ischemic stroke: a systematic review. *Cerebrovascular diseases* **20**, 213-219 (2005).

27. Wunderlich, M.T., Wallesch, C.W. & Goertler, M. Release of glial fibrillary acidic protein is related to the neurovascular status in acute ischemic stroke. *European journal of neurology* **13**, 1118-1123 (2006).

28. Foerch, C.*, et al.* Evaluation of serum S100B as a surrogate marker for long-term outcome and infarct volume in acute middle cerebral artery infarction. *Archives of neurology* **62**, 1130-1134 (2005).

29. Bustamante, A.*, et al.* Blood Biomarkers for the Early Diagnosis of Stroke: The Stroke-Chip Study. *Stroke; a journal of cerebral circulation* **48**, 2419-2425 (2017).

30. Llombart, V.*, et al.* B-type natriuretic peptides help in cardioembolic stroke diagnosis: pooled data meta-analysis. *Stroke; a journal of cerebral circulation* **46**, 1187-1195 (2015).

31. Fonseca, A.C.*, et al.* N-terminal pro-brain natriuretic peptide shows diagnostic accuracy for detecting atrial fibrillation in cryptogenic stroke patients. *International journal of stroke : official journal of the International Stroke Society* **9**, 419-425 (2014).

32. Montaner, J.*, et al.* Etiologic diagnosis of ischemic stroke subtypes with plasma biomarkers. *Stroke; a journal of cerebral circulation* **39**, 2280-2287 (2008).

33. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297 (2004).

34. Mendell, J.T. & Olson, E.N. MicroRNAs in stress signaling and human disease. *Cell* **148**, 1172-1187 (2012).

35. Yates, L.A., Norbury, C.J. & Gilbert, R.J. The long and short of microRNA. *Cell* **153**, 516-519 (2013).

36. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nature reviews. Genetics* **16**, 421-433 (2015).

37. Jimenez-Mateos, E.M.*, et al.* Silencing microRNA-134 produces neuroprotective and prolonged seizure-suppressive effects. *Nature medicine* **18**, 1087-1094 (2012).

38. Kornfeld, J.W.*, et al.* Obesity-induced overexpression of miR-802 impairs glucose metabolism through silencing of Hnf1b. *Nature* **494**, 111-115 (2013).

39. Thum, T.*, et al.* MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature* **456**, 980-984 (2008).

40. Schober, A.*, et al.* MicroRNA-126-5p promotes endothelial proliferation and limits atherosclerosis by suppressing Dlk1. *Nature medicine* **20**, 368-376 (2014).

41. Mitchell, P.S.*, et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 10513-10518 (2008).

42. Cortez, M.A.*, et al.* MicroRNAs in body fluids--the mix of hormones and biomarkers. *Nat Rev Clin Oncol* **8**, 467-477 (2011).

43. Arroyo, J.D.*, et al.* Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 5003-5008 (2011).

44. Vickers, K.C., Palmisano, B.T., Shoucri, B.M., Shamburek, R.D. & Remaley, A.T. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nature cell biology* **13**, 423-433 (2011).

45. Bang, C.*, et al.* Cardiac fibroblast-derived microRNA passenger strand-enriched exosomes mediate cardiomyocyte hypertrophy. *The Journal of clinical investigation* **124**, 2136-2146 (2014).

46. Creemers, E.E., Tijsen, A.J. & Pinto, Y.M. Circulating microRNAs: novel biomarkers and extracellular communicators in cardiovascular disease? *Circulation research* **110**, 483-495 (2012).

47. He, Y.*, et al.* Current State of Circulating MicroRNAs as Cancer Biomarkers. *Clinical chemistry* **61**, 1138-1155 (2015).

48. Valadi, H.*, et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature cell biology* **9**, 654-659 (2007).

49. Laffont, B.*, et al.* Activated platelets can deliver mRNA regulatory Ago2*microRNA complexes to endothelial cells via microparticles. *Blood* **122**, 253-261 (2013).

50. Ong, S.G.*, et al.* Cross talk of combined gene and cell therapy in ischemic heart disease: role of exosomal microRNA transfer. *Circulation* **130**, S60-69 (2014).

51. Tan, K.S.*, et al.* Expression profile of MicroRNAs in young stroke patients. *PloS one* **4**, e7689 (2009).

52. Li, P.*, et al.* Identification of circulating microRNAs as potential biomarkers for detecting acute ischemic stroke. *Cellular and molecular neurobiology* **35**, 433-447 (2015).

53. Monaghan, P.J.*, et al.* Biomarker development targeting unmet clinical needs. *Clinica chimica acta; international journal of clinical chemistry* **460**, 211-219 (2016).

54. Horvath, A.R.*, et al.* From biomarkers to medical tests: the changing landscape of test evaluation. *Clinica chimica acta; international journal of clinical chemistry* **427**, 49-57 (2014).

55. Wardlaw, J.M. & Mielke, O. Early signs of brain infarction at CT: observer reliability and outcome after thrombolytic treatment--systematic review. *Radiology* **235**, 444-453 (2005).

56. Zampetaki, A. & Mayr, M. MicroRNAs in vascular and metabolic disease. *Circulation research* **110**, 508-522 (2012).

57. Ramanujam, D. & Engelhardt, S. Intercellular miRNA Traffic. *Circulation research* **116**, 1726-1728 (2015).

58. Che, P.*, et al.* miR-125a-5p impairs endothelial cell angiogenesis in aging mice via RTEF-1 downregulation. *Aging cell* **13**, 926-934 (2014).

59. Muramatsu, F., Kidoya, H., Naito, H., Sakimoto, S. & Takakura, N. microRNA-125b inhibits tube formation of blood vessels through translational suppression of VE-cadherin. *Oncogene* **32**, 414-421 (2013).

60. McAlexander, M.A., Phillips, M.J. & Witwer, K.W. Comparison of Methods for miRNA Extraction from Plasma and Quantitative Recovery of RNA from Cerebrospinal Fluid. *Frontiers in genetics* **4**, 83 (2013).

61. Mestdagh, P.*, et al.* Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nature methods* **11**, 809-815 (2014).

62. Willeit, P.*, et al.* Circulating microRNAs as novel biomarkers for platelet activation. *Circulation research* **112**, 595-600 (2013).

63. Zampetaki, A.*, et al.* Plasma microRNA profiling reveals loss of endothelial miR-126 and other microRNAs in type 2 diabetes. *Circulation research* **107**, 810-817 (2010).

64. Cohen, L., Hartman, M.R., Amardey-Wellington, A. & Walt, D.R. Digital direct detection of microRNAs using single molecule arrays. *Nucleic acids research* **45**, e137 (2017).

65. McArdle, H.*, et al.* "TORNADO" - Theranostic One-Step RNA Detector; microfluidic disc for the direct detection of microRNA-134 in plasma and cerebrospinal fluid. *Scientific reports* **7**, 1750 (2017).

66. Hofmann, S.*, et al.* Double-Stranded Ligation Assay for the Rapid Multiplex Quantification of MicroRNAs. *Analytical chemistry* **87**, 12104-12111 (2015).

67. Williams, M.R., Stedtfeld, R.D., Stedtfeld, T.M., Tiedje, J.M. & Hashsham, S.A. Quantification of microRNAs directly from body fluids using a base-stacking

isothermal amplification method in a point-of-care device. *Biomed Microdevices* **19**, 45 (2017).

68. Karakas, M. & Zeller, T. A Biomarker Ocular - Circulating MicroRNAs Toward Diagnostics for Acute Ischemic Stroke. *2017* **121**, 905-907 (2017).

69. Ay, H.*, et al.* A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke; a journal of cerebral circulation* **38**, 2979-2984 (2007).

70. Arsava, E.M.*, et al.* The Causative Classification of Stroke system: an international reliability and optimization study. *Neurology* **75**, 1277-1284 (2010).

71. Hopkin, R.J.*, et al.* The management and treatment of children with Fabry disease: A United States-based perspective. *Mol Genet Metab* **117**, 104-113 (2016).

72. Pirmohamed, M.*, et al.* A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med* **369**, 2294-2303 (2013).

73. Kimmel, S.E.*, et al.* A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med* **369**, 2283-2293 (2013).

74. Gage, B.F.*, et al.* Effect of Genotype-Guided Warfarin Dosing on Clinical Events and Anticoagulation Control Among Patients Undergoing Hip or Knee Arthroplasty - The GIFT Randomized Clinical Trial. *JAMA : the journal of the American Medical Association* **318**, 1115-1124 (2017).

75. Azghandi, S.*, et al.* Deficiency of the stroke relevant HDAC9 gene attenuates atherosclerosis in accord with allele-specific effects at 7p21.1. *Stroke; a journal of cerebral circulation* **46**, 197-202 (2015).

76. Strbian, D.*, et al.* Ultra-early intravenous stroke thrombolysis: do all patients benefit similarly? *Stroke; a journal of cerebral circulation* **44**, 2913-2916 (2013).

77. Kunz, A.*, et al.* Functional outcomes of pre-hospital thrombolysis in a mobile stroke treatment unit compared with conventional care: an observational registry study. *Lancet neurology* **15**, 1035-1043 (2016).

78. Fassbender, K.*, et al.* Mobile stroke units for prehospital thrombolysis, triage, and beyond: benefits and challenges. *Lancet neurology* **16**, 227-237 (2017).

79. Meretoja, A.*, et al.* Reducing in-hospital delay to 20 minutes in stroke thrombolysis. *Neurology* **79**, 306-313 (2012).

80. Sharma, K.*, et al.* Cell type- and brain region-resolved mouse brain proteome. *Nature neuroscience* **18**, 1819-1831 (2015).

81. Gerschenfeld, G.*, et al.* Two Paradigms for Endovascular Thrombectomy After Intravenous Thrombolysis for Acute Ischemic Stroke. *JAMA Neurol* **74**, 549-556 (2017).

82. Menon, B.K.*, et al.* Analysis of Workflow and Time to Treatment on Thrombectomy Outcome in the Endovascular Treatment for Small Core and Proximal Occlusion Ischemic Stroke (ESCAPE) Randomized, Controlled Trial. *Circulation* **133**, 2279-2286 (2016).

83. Ahmad, O., Wardlaw, J. & Whiteley, W.N. Correlation of levels of neuronal and glial markers with radiological measures of infarct volume in ischaemic stroke: a systematic review. *Cerebrovascular diseases* **33**, 47-54 (2012).

84. Chehaibi, K., Trabelsi, I., Mahdouani, K. & Slimane, M.N. Correlation of Oxidative Stress Parameters and Inflammatory Markers in Ischemic Stroke Patients. *J Stroke Cerebrovasc Dis* **25**, 2585-2593 (2016).

85. Tarnowski, D.*, et al.* Comparison of diverse platelet activation markers as indicators for left atrial thrombus in atrial fibrillation. *Platelets*, 1-7 (2017).

86. Lavallee, P.C.*, et al.* Circulating markers of endothelial dysfunction and platelet activation in patients with severe symptomatic cerebral small vessel disease. *Cerebrovascular diseases* **36**, 131-138 (2013).

87. Kaudewitz, D.*, et al.* Association of MicroRNAs and YRNAs With Platelet Function. *Circulation research* **118**, 420-432 (2016).

88. Veltkamp, R. & Neurologie, K.L.d.D.G.f. S1-Leitlinie Akuttherapie des ischämischen Schlaganfalls. *Deutsche Gesellschaft für Neurologie, Hrsg. Leitlinien für Diagnostik und Therapie in der Neurologie.* (2012).

89. Nadar, S.K., Lip, G.Y., Lee, K.W. & Blann, A.D. Circulating endothelial cells in acute ischaemic stroke. *Thrombosis and haemostasis* **94**, 707-712 (2005).

90. Damani, S.*, et al.* Characterization of circulating endothelial cells in acute myocardial infarction. *Sci Transl Med* **4**, 126ra133 (2012).

91. Endres, M., Diener, H.C., Behnke, M. & Röther, J. S3-Leitlinie Sekundärprophylaxe ischämischer Schlaganfall und transitorische ischämische Attacke. *Deutsche Gesellschaft für Neurologie, Hrsg. Leitlinien für Diagnostik und Therapie in der Neurologie.* (2015).

92. Hart, R.G.*, et al.* Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet neurology* **13**, 429-438 (2014).

93. Kamel, H. & Healey, J.S. Cardioembolic Stroke. *Circulation research* **120**, 514-526 (2017).

7. Acknowledgements

This work was only accomplished because also many other people contributed with their skills and spirit, worked very hard and believed in my work on this project. Martin Dichgans conceptualized this work and provided an excellent working environment with many other magnificient researchers in house or on campus. Matthias Prestel brought his excellent biochemical and technical knowledge to this project and together, we finetuned many aspects of this work. I want to give special thanks to Veronika Kautzky, Ivelina Stoycheva and Julia Böck, who showed big commitment around the clock in the clinical part of this work. Rainer Malik has been of tremendous help for statistical analysis always open to share his knowledge. During this work I also had the chance to collaborate with Lesca Holdt, Marco Duering, Bernd Northoff, Nick Plesnila, Uta Mamrak, Arthur Liesz, Gemma Llovera, Daniel Teupser, Matthias Klein and Franziska Dorn. In one way or the other, they all have devoted some of their time and energy to support this project and my work.

Danke euch, meinen beiden Eltern, für all die liebevolle Unterstützung und manchmal notwendige Refokussierung meiner Gedanken während all dieser Jahre meiner Ausbildung. Es bedeutet mir sehr viel, dass ich euch auch ein bisschen stolz mache mit meiner Arbeit.

Μελίνα αγάπη μου, σε ευχαριστώ, χωρίς εσένα όλα είναι ενα τίποτα.

8. List of publications

2017

Tiedt S, Prestel M, Malik R, Schieferdecker N, Düring M, Kautzky V, Stoycheva I, Böck J, Northoff BH, Klein M, Dorn F, Krohn K, Teupser D, Liesz A, Plesnila N, Holdt LM, Dichgans M. RNA-Seq Identifies Circulating miR-125a-5p, miR-125b-5p and miR-143-3p as Potential Biomarkers for Acute Ischemic Stroke. Circ Res. 2017 Sep 29;121(8):970-980. *(Impact factor (IF) 2017: 14.0)*

2016

Pulit SL, …, **Tiedt S**, …, Worrall BB; NINDS Stroke Genetics Network (SiGN); International Stroke Genetics Consortium (ISGC). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. Lancet Neurol. 2016 Feb;15(2):174-184. *(IF 2016: 26.3)*

Meissner L, Gallozzi M, Balbi M, Schwarzmaier S, **Tiedt S**, Terpolilli NA, Plesnila N. Temporal Profile of MicroRNA Expression in Contused Cortex after Traumatic Brain Injury in Mice. J Neurotrauma. 2016 Apr 15;33(8):713-20. *(IF 2016: 5.2)*

Kunisch R, Guder P, Schinke K, Nörenberg D, Ruf VC, Alig S, Bauer HJ, Kirchner SK, Kruger S, Noerenberg D, Singer K, **Tiedt S**, Weckbach L, Wypior G, Angstwurm M. Reforming the Surgical Section of the Practical Year at Ludwig-Maximilians-University Munich. Zentralbl Chir. 2016 Jun;141(3):310-4. *(IF 2016: 1.0)*

2014

McArdle PF, …, **Tiedt S**, …, Worrall BB; NINDS SiGN Study. Agreement between TOAST and CCS ischemic stroke classification: the NINDS SiGN study. Neurology. 2014 Oct 28;83(18):1653-60. *(IF 2014: 8.3)*

Levin J*, **Tiedt S*,** Arzberger T, Biskup S, Schuberth M, Stenglein-Krapf G, Kreth FW, Högen T, la Fougère C, Linn J, van der Knaap MS, Giese A, Kretzschmar HA, Danek A. Diffuse leukoencephalopathy with spheroids: biopsy findings and a novel mutation. Clin Neurol Neurosurg. 2014 Jul;122:113-5. * These two authors contributed equally. *(IF 2013: 1.3)*

2013

Sirko S, Behrendt G, Johansson PA, Tripathi P, Costa M, Bek S, Heinrich C, **Tiedt S**, Colak D, Dichgans M, Fischer IR, Plesnila N, Staufenbiel M, Haass C, Snapyan M, Saghatelyan A, Tsai LH, Fischer A, Grobe K, Dimou L, Götz M. Reactive glia in the injured brain acquire stem cell properties in response to sonic hedgehog. Cell Stem Cell. 2013 Apr 4;12(4):426-39. *(IF 2013: 22.2)*

Pinilla S*, Weckbach LT*, Alig SK, Bauer H, Noerenberg D, Singer K, **Tiedt S**. Blogging Medical Students: A Qualitative Analysis. GMS Z Med Ausbild. 2013;30(1):Doc9. * These two authors contributed equally. *(IF: not available)*

2010

Heinrich C, Blum R, Gascon S, Masserdotti G, Tripathi P, Sanchez R, **Tiedt S**, Schröder T, Götz M and Berninger B. Directing Astroglia from the cerebral cortex into subtype specific functional neurons. PLoS Biol 2010 May 18; 8(5):e1000373. *(IF 2010: 12.5)*

9. Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation "**The role of blood-based biomarkers in ischemic stroke**" selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schriftum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter der Bezeichnung der Fundstelle einzeln nachgewiesen habe.

München, 23.10.2017

_____

Steffen Tiedt

10. Declaration of author contributions

Authors contributed to the manuscripts as follows:

1. **Tiedt S**, Prestel M, Malik R, Schieferdecker N, Duering M, Kautzky V, Stoycheva I, Böck J, Northoff B, Klein M, Dorn F, Krohn K, Teupser D, Liesz A, Plesnila N, Holdt LM, Dichgans M. **RNA-seq identifies circulating miR-125a-5p, miR-125b-5p and miR-143-3p as potential biomarkers for acute ischemic stroke.** *ST, MP and M. Dichgans designed the experiments. ST, NS and KK performed the experiments. ST, VK, IS, JB and MK recruited patients and healthy control subjects. ST, RM, M. Duering, BN and FD analysed the experiments. ST, MP, DT, AL, NP, LMH and M. Dichgans interpreted the results. ST and M. Dichgans wrote the manuscript. MP, RM, M. Duering, AL, NP and LMH edited the manuscript.*

2. McArdle PF, Kittner SJ, Ay H, Brown RD Jr, Meschia JF, Rundek T, Wassertheil-Smoller S, Woo D, Andsberg G, Biffi A, Brenner DA, Cole JW, Corriveau R, de Bakker PI, Delavaran H, Dichgans M, Grewal RP, Gwinn K, Huq M, Jern C, Jimenez-Conde J, Jood K, Kaplan RC, Katschnig P, Katsnelson M, Labovitz DL, Lemmens R, Li L, Lindgren A, Markus HS, Peddareddygari LR, Pedersén A, Pera J, Redfors P, Roquer J, Rosand J, Rost NS, Rothwell PM, Sacco RL, Sharma P, Slowik A, Sudlow C, Thijs V, **Tiedt S**, Valenti R, Worrall BB; NINDS SiGN Study. **Agreement between TOAST and CCS ischemic stroke classification: The NINDS SiGN Study.** *ST classified IS patients according to their etiology.*

3. Pulit SL, …, **Tiedt S**, …, Worrall BB; NINDS Stroke Genetics Network (SiGN) and International Stroke Genetics Consortium (ISGC). **Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study.** *ST collected blood samples and classified IS patients according to their etiology.*

Herewith, I confirm the contributions to the manuscripts.

München, 23.10.2017

_____	_____
Steffen Tiedt	Prof. Dr. Martin Dichgans
	(supervisor)