Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke

Joan Montaner[1,2,3,10]✉, *Laura Ramiro*[1,10], *Alba Simats*[1], *Steffen Tiedt*[4], *Konstantinos Makris*[5], *Glen C. Jickling*[6], *Stephanie Debette*[7,8], *Jean-Charles Sanchez*[9] *and Alejandro Bustamante*[1]

Abstract | Despite many years of research, no biomarkers for stroke are available to use in clinical practice. Progress in high-throughput technologies has provided new opportunities to understand the pathophysiology of this complex disease, and these studies have generated large amounts of data and information at different molecular levels. The integration of these multi-omics data means that thousands of proteins (proteomics), genes (genomics), RNAs (transcriptomics) and metabolites (metabolomics) can be studied simultaneously, revealing interaction networks between the molecular levels. Integrated analysis of multi-omics data will provide useful insight into stroke pathogenesis, identification of therapeutic targets and biomarker discovery. In this Review, we detail current knowledge on the pathology of stroke and the current status of biomarker research in stroke. We summarize how proteomics, metabolomics, transcriptomics and genomics are all contributing to the identification of new candidate biomarkers that could be developed and used in clinical stroke management.

Ischaemic core
The region of the brain with the most severe blood flow deficits (blood flow below 10–25%), resulting in rapid progression of cell death.

✉e-mail: joan.montaner@vhir.org

https://doi.org/10.1038/s41582-020-0350-6

Neurovascular diseases are increasingly common owing to an increasing burden of vascular risk factors and to ageing populations. According to the American Heart Association, on average, someone experiences a stroke every 40 s in the United States, and stroke prevalence in the United States will increase by up to 20.5% between 2012 and 2030. Ischaemic stroke is the most common type, accounting for 87% of all strokes[1]. The only approved therapy is recanalization of the occluded artery via administration of intravenous recombinant tissue-plasminogen activator (tPA), followed by endovascular mechanical thrombectomy in the case of large-vessel occlusions.

A better understanding of the pathophysiology of neurovascular diseases will improve preventive, diagnostic, therapeutic and reparative strategies. Identification and monitoring of molecular biomarkers that are present during the natural course of ischaemic stroke would aid stroke diagnosis, prognosis and development of therapeutic strategies[2]. Any biomarker that increases the chance that a patient with stroke will be treated within 60 min of onset could be as beneficial as, but much more affordable than, mobile stroke units, which increase the odds of survival, of regaining independence and of becoming asymptomatic[3].

In this Review, we provide an overview of how multi-omics techniques are contributing to the discovery of biomarkers for diagnosis and prognosis in ischaemic stroke, and how they are helping to identify molecular targets for therapeutic interventions (FIG. 1). We consider the progress made in genomics, transcriptomics, proteomics and metabolomics, and discuss how bringing data from these techniques together through integromics and systems biology will move our understanding and management of stroke forwards. We focus on ischaemic stroke, as the current status of biomarkers for haemorrhagic stroke has been reviewed elsewhere[4].

Pathophysiology of ischaemic stroke

In acute ischaemic stroke, a sudden decline in cerebral blood flow causes a total or partial reduction in the oxygen and glucose supply to neurons and other brain cells. As a consequence, multiple physiological, biochemical and molecular mechanisms are triggered[5], which ultimately lead to extensive cell death and alter basic neurological functions in the ischaemic core. In the penumbra — the area that surrounds the ischaemic core — brain cells remain metabolically active and structurally intact for longer, so are salvageable with immediate therapy[6].

Dying brain cells release danger signals — so-called damage-associated molecular patterns (DAMPs) — that stimulate post-stroke inflammation (reviewed in detail elsewhere[7–9]). DAMPs promote expression of pro-

Key points

- Biomarkers of stroke could improve diagnosis and management, but standardization or harmonization of procedures is needed before translation of biomarkers to clinical practice to ensure results are comparable and reliable.
- Studies of the proteome of the brain, cerebrospinal fluid and brain extracellular fluid after ischaemic stroke have led to identification of candidate biomarkers for stroke management.
- Most studies of stroke genetics have focused on common or low-frequency single-nucleotide polymorphisms; other types of variation, such as rare single-nucleotide variants or structural variations, have been insufficiently explored.
- Changes in RNA levels in stroke have the potential to aid stroke diagnosis and provide insight into stroke aetiology.
- Circulating metabolites provide information about local and systemic events after stroke, and therefore could serve as biomarkers of stroke and for differentiation of major stroke aetiologies.
- Integrated analysis of data obtained with different omics approaches will enable implementation of biomarkers at several stages in the stroke care pathway, with the potential to transform stroke management.

inflammatory cytokines, chemokines and adhesion molecules by resident immune cells, which help circulating leukocytes to infiltrate the infarcted brain[10–12]. These pro-inflammatory mediators also contribute to disruption of the blood–brain barrier (BBB)[13], increasing its permeability and thereby further facilitating accumulation of immune cells in the injured brain[14]. This disruption of the BBB breaks the immune privilege of the brain and exposes neuronal antigens to the periphery, further stimulating the inflammatory response. In addition, DAMPs and pro-inflammatory cytokines activate primary and secondary lymphoid organs, resulting in a systemic inflammatory response syndrome. Thus, post-stroke immunological changes are not limited to the brain but also occur in various peripheral organs, including the blood, bone marrow, spleen and gut[15].

However, the inflammatory response after stroke is self-limiting. Excessive amounts of pro-inflammatory mediators also activate the sympathetic nervous system and the hypothalamic–pituitary–adrenal axis[16]. These systems release stress hormones, such as glucocorticoids and catecholamines[17], which promote apoptosis of immune cells, increase production of anti-inflammatory cytokines, such as transforming growth factor-β (TGFβ), and inhibit production of pro-inflammatory mediators, such as IL-1, IL-8 and tumour necrosis factor (TNF)[18]. As a consequence of this immunosuppression, however, the risk of infection after stroke onset is substantially increased[19,20]; complications from respiratory or urinary tract infections occur in ~30% of all patients with ischaemic stroke[21].

Among the thousands of molecules that are altered during these processes, some might be valuable indicators of disease and could help with diagnosis and prognosis of ischaemic stroke. Pathologically deregulated molecules could also be meaningful therapeutic targets that could address the urgent need for treatments. High-throughput omics technologies are already identifying some molecules with value, and these are discussed in detail below.

Stroke biomarkers — current status

Although blood biomarkers have been used in the management of other vascular conditions for many years, no biomarker is specifically used for stroke management. Some candidate markers of specific clinical indications during the time course of stroke have emerged thanks to the gradual build-up of knowledge and could be implemented in clinics in the near future.

Acute ischaemic stroke. Research into biomarkers in acute stroke has focused on two aims. The first is to improve or accelerate diagnosis of stroke. The second is to improve clinical decision-making and facilitate use of personalized therapy.

Currently, stroke is diagnosed by clinical assessment and neuroimaging. A faster, accurate technique for stroke diagnosis, such as a blood test, is essential to enable earlier treatment. In previous studies, the utility of single biomarkers or biomarker panels in this context have been assessed, but the accuracies achieved have been far below what is required in clinical practice[22–24]. A diagnostic biomarker of transient ischaemic attack would also be of great value because many patients are already asymptomatic when attended by doctors, so diagnosis mainly relies on clinical history[25].

A bedside test that not only aids diagnosis of stroke but also helps to differentiate between stroke subtypes — particularly to distinguish ischaemic stroke from intracerebral haemorrhage (ICH) — would also improve management. In-hospital CT is 100% sensitive for acute bleeds but can only be done in hospital, whereas pre-hospital thrombolysis for ischaemic stroke is feasible[26] and intracerebral bleeding needs to be ruled out in the pre-hospital setting. One candidate biomarker in this context is glial fibrillary acidic protein (GFAP), a highly brain-specific intermediate filament protein that is rapidly released into the bloodstream after ICH. A meta-analysis has shown that measurement of GFAP in the blood could discriminate between ICH and ischaemic stroke, as circulating levels of GFAP are higher in ICH[27,28]. Another promising biomarker for distinguishing between stroke subtypes is retinol-binding protein 4 (RBP4). The combination of these two biomarkers improves discrimination[29].

Biomarkers that predict responses to therapy are also increasingly important because different patients

Author addresses

[1]Neurovascular Research Laboratory, Vall d'Hebron Institute of Research, Universitat Autònoma de Barcelona, Barcelona, Spain.
[2]Institute of Biomedicine of Seville, IBiS/Hospital Universitario Virgen del Rocío/CSIC/University of Seville, Seville, Spain.
[3]Department of Neurology, Hospital Universitario Virgen Macarena, Seville, Spain.
[4]Institute for Stroke and Dementia Research, University Hospital, LMU Munich, Munich, Germany.
[5]Clinical Biochemistry Department, KAT General Hospital, Kifissia, Athens, Greece.
[6]Department of Neurology, University of Alberta, Edmonton, AB, Canada.
[7]University of Bordeaux, Inserm, Bordeaux Population Health Research Center, team VINTAGE, UMR 1219, Bordeaux, France.
[8]Department of Neurology, CHU de Bordeaux, Bordeaux, France.
[9]Translational Biomarker Group, Faculty of Medicine, University of Geneva, Geneva, Switzerland.
[10]These authors contributed equally: Joan Montaner, Laura Ramiro.

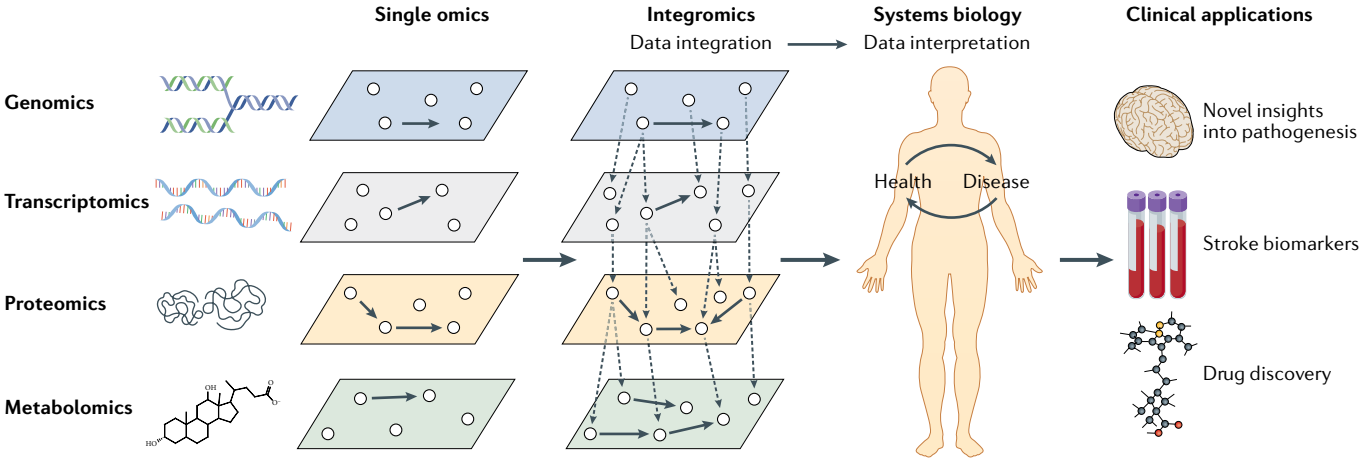

Fig. 1 | **Schematic representation of a multi-omics approach to the study of stroke.** Single omics data can be combined through integromics and systems biology. This multidisciplinary approach contributes to a better understanding of the mechanisms involved in stroke and facilitates the development of biomarkers and therapeutics.

respond differently to intravenous thrombolysis and mechanical reperfusion. Specifically, biomarkers that predict haemorrhagic transformation or that identify patients for whom the chance of recanalization after thrombolysis is low would be valuable. Several studies have demonstrated an association between haemorrhagic transformation and increased levels of circulating matrix metalloproteinase 9 (MMP9), which plays a key role in microvascular and BBB integrity[30–32]. A meta-analysis that included 12 clinical studies showed that measurement of MMP9 had a high sensitivity (85%) but suboptimal (79%) specificity for haemorrhagic transformation after acute ischaemic stroke[33]. Other candidate biomarkers for haemorrhagic transformation include fibronectin[34] and the semicarbazide-sensitive amine oxidase activity of vascular adhesion protein 1 (VAP1)[35].

In relation to thrombolysis, pretreatment levels of fibrinolytic inhibitors, such as low plasminogen activator inhibitor 1 (PAI1) or high thrombin activatable fibrinolysis inhibitor (TAFI) levels, have been associated with a poor response to intravenous tPA[36]. Low circulating levels of a disintegrin and metalloproteinase with thrombospondin motifs 13 (ADAMTS13) have also been associated with a poor response after recanalization therapies, both after intravenous tPA and mechanical thrombectomy[37].

Biomarkers that predict the response to intravenous tPA could be used at the pre-hospital stage or in centres where endovascular therapy is unavailable to facilitate the transfer of patients to comprehensive stroke centres when poor recanalization with intravenous treatment is anticipated. A test in which different biomarkers are combined to predict the efficacy and safety of intravenous tPA for an individual would be ideal for clinical practice.

Beyond acute ischaemic stroke. Post-stroke outcomes after several months differ between patients even when their baseline characteristics are similar. Many molecular markers have been associated with these outcomes. For example, levels of inflammatory biomarkers such as IL-6 and copeptin (the C-terminal portion of pro-vasopressin) have greater predictive value for outcomes than do clinical variables[38]. However, in addition to facilitating predictions about overall outcomes, blood biomarkers could provide information about specific causes of complications that lead to poor outcomes[39].

For example, some markers have been associated with outcomes from stroke-associated infections. Monocytic expression of the human leukocyte antigen gene *HLA-DR* in the blood is an independent predictor of an individual's ability to overcome post-stroke infections[40]. In addition, some evidence suggests that expression of procalcitonin, which is used as a diagnostic marker of bacterial infections in hospitalized patients, is positively associated with post-stroke infections[41], although another study (the STRAWINSKI trial) indicated that procalcitonin-guided antibiotic therapy was not superior to conventional management[42]. Another promising candidate marker in this context is serum amyloid A (SAA), an acute-phase reactant that had a high sensitivity for infections in patients with stroke in one study[43].

Among patients with ischaemic stroke, 25–30% are discharged without a specific aetiological diagnosis of stroke despite an extensive work-up. Paroxysmal atrial fibrillation has been proposed as the underlying cause in a large percentage of these patients, especially those with embolic stroke of undetermined source[44]. Possible biomarkers of cardioembolic stroke are B-type natriuretic peptides (BNP and NT-proBNP), blood levels of which are increased in this subtype[45,46]. Moreover, high levels of natriuretic peptides have been associated with a cardioembolic source among patients who were discharged with an undetermined cause of stroke[47]. Clinical trials to investigate whether anticoagulants can prevent further strokes in patients who have had a cryptogenic stroke and have high levels of natriuretic peptides are warranted, and one is already in progress[48]. Indeed, this potential biomarker of a stroke subtype looks to be the closest to clinical use.

The use of aetiology-specific biomarkers could improve primary prevention strategies as well as

Embolic stroke of undetermined source
Ischaemic stroke with an unknown origin; these strokes are non-lacunar and non-atherosclerotic strokes of an undetermined embolic source.

secondary prevention. The role of natriuretic peptides in screening for atrial fibrillation among the general population is currently under investigation in a population-based study[49]. Biomarkers could, therefore, aid not just stroke management and secondary prevention but also personalized medicine for stroke even before it occurs. This possibility could dramatically alter the path of a patient through stroke care (TABLE 1).

Standardization of biomarker assays. Although several biomarkers might be close to clinical readiness, their application in clinical practice will require further research owing to sources of variability at pre-analytical and analytical stages. Pre-analytical sources include sample storage conditions and time, freeze–thaw cycles, types of biological sample (for example, serum, plasma and other fluids) and biological variation of the biomarker in health and disease. Analytical variability is introduced by factors connected to the analytical method, such as the assay design (sandwich or competitive), antibodies (used as reagents) or differences in calibrators (usually recombinant proteins or synthetic peptides used to calibrate immunoassays)[50,51]. Standardization of measurements in plasma is particularly difficult owing to the heterogeneity of this fluid. Nevertheless, the widespread use and simple acquisition of plasma samples in laboratory diagnostics and healthcare mean that plasma components are ideal targets for standardization, or at least harmonization[52]. The use of non-standardized assays in clinical research can lead to inaccurate conclusions about the value of biomarkers and, ultimately, lead to wrong clinical decisions. Key to addressing this issue will be recalibration of old measurements from clinical and epidemiological studies to gold-standard reference measurements once standardization or harmonization has been achieved. This approach will enable meta-analyses that will produce robust conclusions to support clinical decisions. Moreover, standardization of clinical data could enable data analysis across trials; the larger sample sizes created by this possibility would provide sufficient power to study subgroups, which cannot be explored in single trials.

Multi-omics for biomarker discovery

Most biomarkers of stroke discussed above have been identified on the basis of underlying pathophysiological processes, such as inflammation, immune system responses, apoptosis, coagulation, fibrinolysis, tissue remodelling and heart damage. However, advances in technology are enabling use of high-throughput techniques based on large screening processes to avoid selection bias and

Table 1 | **Current standard of care in stroke prevention and treatment compared with future biomarker-based stroke care**

Disease stage	Stroke care in 2020	Future, biomarker-based stroke care	Useful biomarkers
Health	Management of vascular risk factors according to guidelines; opportunistic screening for atrial fibrillation and atherosclerosis	Management of vascular risk factors guided by biomarker-based risk stratification; biomarker-based screening for atrial fibrillation	Lp-PLA2 and ADAMTS13 (risk stratification); natriuretic peptides (screening for atrial fibrillation)
Silent brain infarcts	Management of vascular risk factors according to guidelines; opportunistic screening for atrial fibrillation and atherosclerosis	Biomarkers for detection of subclinical vascular brain injuries; optimization of primary prevention	Lp-PLA2; natriuretic peptides (prediction of silent brain infarcts)
Acute stroke	Pre-hospital diagnosis: differentiation between ischaemic stroke and ICH with CT; detection of LVO with CTA; assessment of penumbra with CTP	Pre-hospital diagnosis: differentiation between ischaemic stroke, ICH and mimics with biomarker panels; optimization of patient transfer according to biomarker-based diagnosis	RBP4 and GFAP (stroke diagnosis, pre-hospital treatment); MMP9 (prediction of haemorrhagic transformation); ADAMTS13 (prediction of response to thrombolysis or mechanical thrombectomy)
	Pre-hospital treatment: none	Pre-hospital treatment: intravenous tPA in selected patients with ischaemic stroke identified with biomarker panels; intensive blood pressure lowering in ICH identified with biomarkers	
	Acute treatment: intravenous tPA or mechanical thrombectomy for ischaemic stroke; blood pressure management for ICH	Acute treatment: intravenous tPA or mechanical thrombectomy for ischaemic stroke; personalized strategies depending on biomarkers	
Subacute stroke	Stroke unit admission; prevention of complications by following established protocols (cardiac monitoring, blood pressure management, dysphagia assessment); aetiological work-up, including carotid ultrasounds, conventional blood tests and long-term cardiac monitoring, if aetiology is undetermined	Stroke unit admission, neurological and functional outcome prediction; prevention and early treatment of complications based on blood biomarkers; biomarker-based aetiological work-up for long-term monitoring	Copeptin, IL-6 (long-term prognosis); SAA and mHLA-DR (stroke-associated infections); hs-Tn (cardiac complications); natriuretic peptides (cardioembolic aetiology)
Secondary prevention	Based on stroke aetiology, but management of general vascular risk factors and antiplatelet treatment in cryptogenic stroke	Based on stroke aetiology, but based on biomarkers in cryptogenic stroke	Natriuretic peptides (cardioembolic aetiology)

ADAMTS13, a disintegrin and metalloproteinase with thrombospondin motifs 13; CTA, CT angiography; CTP, CT perfusion; GFAP, glial fibrillary acid protein; hs-Tn, high-sensitivity troponin; ICH, intracerebral haemorrhage; Lp-PLA2, lipoprotein-associated phospholipase A2; LVO, large-vessel occlusion; mHLA-DR, monocyte HLA-DR; MMP9, matrix metalloproteinase 9; RBP4, retinol-binding protein 4; SAA, serum amyloid A; tPA, tissue-plasminogen activator.

Table 2 | **Potential protein biomarkers in stroke identified in proteomics studies**

Protein	Source samples	Evidence	Refs
Stroke diagnosis			
MMPs (MMP1, MMP2, MMP3, MMP8, MMP9, MMP10, MMP13)	Brain homogenate, neuron and vessel microdissection	All studied MMPs were upregulated in infarcted tissue compared with control areas	54
TIMPs (TIMP1, TIMP2)			
GSTP1	Cerebral microdialysate and blood	Circulating GSTP1, PRDX1 and S100B upregulated in patients with ischaemic stroke	55
PRDX1			
S100B			
H-FABP	CSF and blood	FABP elevated in CSF after stroke; circulating H-FABP elevated after ischaemic stroke	63
CMPK	Rat CSF and human blood	Circulating levels of CKB and CMPK higher in patients with ischaemic stroke than in controls during acute phase (<6 h after symptom onset)	65
CKB			
Stroke prognosis			
Gelsolin	Brain homogenate and blood	Higher baseline levels of gelsolin and cystatin A and lower levels of DRP2 associated with poorer outcomes	55,56
DRP2			
Cystatin A			
SAHH2	Neuron and vessel microdissection and blood	Lower circulating levels of SAHH2 associated with greater neurological improvement at 24 h and 48 h after onset in ischaemic stroke	58
CaMK2B	Rat CSF and human blood	Higher baseline circulating levels of CaMK2B and CMPK associated with poorer functional outcomes 3 months after symptom onset in ischaemic stroke	65
CMPK			

CaMK2B, calmodulin-dependent protein kinase II subunit-β; CKB, creatine kinase B-type; CMPK, cytidine monophosphate kinase; CSF, cerebrospinal fluid; DRP2, dihydropyrimidinase-related protein 2; FABP, fatty acid-binding protein; GSTP1, glutathione S-transferase P1; H-FABP, heart FABP; MMP, matrix metalloproteinase; PRDX1, peroxiredoxin 1; SAHH2, adenosylhomocysteinase 2; TIMP, tissue inhibitor of metalloproteinases.

are generating extensive lists of molecules for evaluation as biomarkers.

Among these omics-based approaches, proteomics (the study of the entire set of proteins) remains the most used tool for discovering new biomarkers[53]. However, genomics, transcriptomics and metabolomics are increasingly generating promising results, and these multi-omics approaches could be complementary, especially if multiple omics approaches can be applied to the same patient.

Proteomics

Technological developments have enabled researchers to move from quantification of individual proteins to simultaneous identification of thousands of proteins. In clinical stroke research, these rapidly progressing proteomics techniques have been used to better understand the pathophysiology of the disease and thereby identify many potential biomarkers (TABLE 2).

Proteomics in the ischaemic brain. In the first study of brain proteomics in patients with stroke, published in 2009, laser microdissection was used to investigate MMP expression profiles in the neurons and vasculature of the ischaemic brain[54]. This study demonstrated that neurons were an important source of MMP10 after stroke, whereas vascular production of MMP9 and tissue inhibitor of metalloproteinases 2 (TIMP2) was high.

Subsequent rapid development of proteomics techniques enabled analysis of the whole brain proteome after ischaemic stroke with a 2D differential gel electrophoresis-based approach[55]. Actin levels were decreased the most in the infarct core in comparison with corresponding contralateral areas, whereas albumin levels were increased the most. Interestingly, the protein expression profile of the penumbral region was intermediate between those of the ischaemic core and the contralateral regions, reinforcing the concept that penumbral tissue is salvageable.

Further improvements in proteomics technologies enabled identification of previously undetected peptides, and by 2013, 51 proteins had been identified as differentially expressed in the infarcted brain after ischaemic stroke[56]. High levels of GFAP, which is related to astrogliosis, have been observed in the infarct core, and evidence suggests that the presence of GFAP in the bloodstream after stroke reflects brain damage and could be used as a test in the clinic when kits for its easy detection are available. Other lesser-known proteins, such as *N*-ethylmaleimide-sensitive factor (NSF) ATPase, which is critical for membrane trafficking in neurons, could become targets to avoid ischaemia–reperfusion injury[57]. Subsequently, circulating levels of gelsolin, dihydropyrimidinase-related protein 2 (DRP2) and cystatin A were discovered to be independent predictors of poor outcome and remain promising as blood biomarkers for stroke prognosis[56].

In a study published in 2018, the latest advances in proteomics enabled evaluation of how each component of the neurovascular unit responds to an ischaemic

event at the protein level[58]. Laser microdissection was used to isolate neuronal and BBB components from the ischaemic core and contralateral areas, and these components were quantified. This study provided the first descriptions of the proteome of neurons and the BBB compartment after ischaemic stroke, increasing our knowledge of the molecular mechanisms involved in stroke in different brain structures. From this analysis, neuronal-specific adenosylhomocysteinase 2 (SAHH2) emerged as a potential blood biomarker that could be used for early stroke prognosis — low circulating levels of SAHH2 were associated with neurological improvements after ischaemic stroke. However, whereas the sensitivity of SAHH2 for the discrimination of stroke outcomes was high (89%), specificity was moderate (58%). Consequently, a combination of SAHH2 with other prognostic biomarkers might be most helpful in the evaluation of clinical stroke outcomes.

Proteomics in body fluids after stroke. In addition to studies of brain proteomics, protein expression patterns in accessible bodily fluids have also been studied to better understand the pathophysiology of ischaemic stroke. In 2011 — a few months after the human brain proteome was described for the first time[55] — the brain extracellular fluid proteome of patients with stroke was also first published[59]. Human cerebral microdialysis enabled in vivo monitoring of changes in the composition of the brain extracellular fluid. Analysis of extracellular fluid samples from the ischaemic core, the penumbra and the contralateral hemisphere revealed 53 proteins that were highly expressed in the ischaemic core and the penumbra when compared with corresponding regions in the contralateral hemisphere. Of these proteins, glutathione *S*-transferase P1 (GSTP1), peroxiredoxin 1 (PRDX1) and protein S100B were proposed as blood biomarkers of stroke because circulating levels of these proteins were also higher after stroke than in healthy controls[59]. PRDX1 and GSTP1 interact and have been implicated in similar redox protective mechanisms, so further study of these proteins in cerebrovascular diseases is highly justified.

Attempts have also been made to study the cerebrospinal fluid (CSF) proteome, but CSF collection is invasive and is contraindicated with intravenous thrombolytic therapies as it can increase the risk of spinal and subdural cerebral haematoma, herniation syndrome and infection after stroke[60,61]. Most studies of the CSF in stroke have therefore focused on post-mortem CSF samples, assuming the post-mortem status to be a model of 'massive brain injury'. In one pilot study, 2D gel electrophoresis separation and mass spectrometry of post-mortem CSF samples showed that CSF levels of fatty acid-binding protein (FABP) were high after death[62]. In another study, post-mortem CSF protein expression was compared with that in healthy individuals, and 13 proteins were found to be differentially expressed after death, indicating that these proteins could be biomarkers of brain damage[63]. Subsequently, a quantitative proteomics approach identified 78 proteins that were increased in post-mortem CSF samples compared with ante-mortem CSF samples. Some of these proteins, such as protein/nucleic acid deglycase DJ-1 (also known as PARK7), S100B and GFAP, have previously been identified as biomarkers of brain damage, thus confirming the utility of post-mortem CSF as a valid model of brain insult[64].

Proteomics studies in humans with stroke can have limitations, such as the time between symptom onset and sample collection in brain studies or contraindications for sample collection in CSF studies, and experimental stroke models are often required to study stroke-specific proteomics changes. For example, the CSF proteome in the hyper-acute phase of stroke has been described in rats using an aptamer-based proteomics assay (SOMAscan) of CSF samples obtained before and very soon after middle cerebral artery occlusion[65]. This analysis revealed that cerebral ischaemia acutely increased levels of 716 proteins in the CSF. Of these, uridine monophosphate–cytidine monophosphate kinase (CMPK), creatine kinase B-type (CKB) and calcium–calmodulin-dependent protein kinase II subunit-α (CaMK2A), subunit-β (CaMK2B) and subunit-δ (CaMK2D) showed promise as biomarkers of ischaemic stroke. On this basis, these proteins were studied in blood samples from patients with ischaemic stroke (taken <6 h after symptom onset) and controls. CKB and CMPK showed potential as diagnostic biomarkers for stroke, and CaMK2B and CMPK showed promise as biomarkers of functional stroke outcome[65].

Limitations and future directions. Tremendous progress in neurovascular proteomics has brought us closer to finding diagnostic, prognostic and treatment biomarkers of ischaemic stroke. However, none of the potential biomarkers identified has yet been translated into an approved clinical tool. One of the main reasons for this translational gap is that the proteomics studies were based on the assumption that 'one size fits all'. However, this assumption is flawed because patients respond differently to treatments as a result of several factors, including genetics and individual proteomics profiles. The hope is that the field of biomarkers will move towards precision medicine, in which variability between individuals is taken into account when choosing the right treatment.

In this context, proteomics can identify protein signatures from thousands of proteins and their associated post-translational modifications. This makes it an ideal strategy for quick discovery of new biomarkers with high precision and predictive power. Patient-centric protein biomarkers will enable better stratification of patients for administration of appropriate treatment at the appropriate dose and at the best time. However, the number of patients required for large proteomics studies and the high cost of the equipment needed are obstacles to rapid clinical implementation.

Another problem that has hindered translation of stroke biomarkers into clinical practice is the complexity of the blood proteome. The total range of protein concentrations is 12 orders of magnitude, making it difficult to identify specific cerebrovascular biomarkers that are present at concentrations from picograms to nanograms per millilitre[66]. One emerging solution to

Box 1 | **Stroke genetic associations for development of therapeutic strategies**

The identification of genetic associations with stroke can contribute to discovery and development of novel therapeutic targets and strategies in several ways:

- By deciphering the molecular mechanisms that underlie stroke risk (and outcome), which informs and accelerates the development of novel therapies by:
  - Identifying genes that are implicated in the occurrence of stroke and its severity and that could be tested as biotargets. This approach requires extensive bioinformatics analysis to identify the most likely causal gene and variant in the associated locus[72] followed by experimental explorations[91,173]. Bioinformatics approaches can also be used to assess the druggability of a putative biotarget.
  - Revealing the potential for drug repositioning[174] through use of bioinformatics tools to determine the enrichment of stroke-associated genes in known drug targets[175]. For example, the MEGASTROKE study revealed enrichment of stroke-associated genes among anti-thrombotic drug target genes, including *PDE3A*, which encodes the target for the antiplatelet drug cilostazol that is widely used for stroke prevention in Asian countries[72].
  - Providing genetic support for drug effects using genetic variants that mimic the effect of a drug (for example, variants that are associated with increased or reduced expression of the drug target gene)[159]. Clinical trials of drugs that have this kind of genetic support are estimated to be twice as successful as those without such genetic support[161].
- By facilitating the classification of stroke into more homogeneous subtypes that might have distinct responses to specific treatments. Molecular signatures of stroke subtypes will be refined by combining genetic associations with other circulating biomarkers[111].
- By identifying individuals who have a high risk of stroke so that they can be targeted for early, intensive preventive interventions[96].

this problem, other than depletion of the most abundant blood proteins, is an antibody-based strategy to enrich cerebrovascular proteins. Within this strategy, promising targets to mediate this enrichment are extracellular vesicles, including exosomes and microparticles. The protein content of extracellular vesicles is specific to their cell of origin, so those that derive from brain endothelial cells are most likely to be good sources of protein biomarkers associated with pathophysiological changes in the brain vessels before, during and after a stroke[67,68].

Another important limitation of proteomics studies is that, although proteomics can detect tens of differentially expressed proteins in a single experiment in a small number of patients, usually only one potential biomarker is further verified and validated in large cohorts of patients. This strategy leads to a single protein diagnostic tool intended for all patients and does not consider the heterogeneity of the patient population. As a consequence, these individual markers are not sufficiently specific or sensitive to be used as in vitro diagnostic tools. To address this issue, several markers discovered by proteomics could be combined to improve clinical performance and patient stratification. In theory, such biomarker panels can combine proteins with clinical parameters, such as age, sex and National Institutes of Health Stroke Scale (NIHSS) score. Unfortunately, few panels that have been developed[69] have been translated into practice, mainly owing to the complexity of establishing a single rule for classification of individuals into a specific group for several parameters, the need for very large clinical studies for cross-validation and the risk of overfitting the performance of the panel due to the high dimensionality of the data (when the number of features exceeds the number of observations; for example, thousands of proteins in a single patient). New statistical strategies and standards need to be implemented at the proteomics screening phase for individual patients to accelerate the translation of panels into precise stroke medicine.

Overfitting
A phenomenon that occurs when a statistical model describes random error or noise instead of the underlying relationship.

Stroke genomics

Unravelling the genomics of stroke has proven more complex than for other common vascular and neurological diseases owing to the heterogeneity of stroke. Important progress has been made with multi-ancestry collaborative efforts in population-based and clinic-based studies. This work has shed new light on the pathophysiology of stroke and opens new avenues for novel therapeutic approaches (BOX 1).

Risk of stroke. In the vast majority of cases, genetic risk variants contribute to a multifactorial predisposition to stroke and each genetic variation is responsible for only modest increases in the risk. Widespread use of high-throughput genotyping in the past decade and the creation of large international consortia, such as the International Stroke Genetics Consortium (ISGC)[70] and the Cohorts of Heart and Ageing Research in Genomic Epidemiology (CHARGE) consortium[71], have led to substantial progress in the discovery of genes that underlie complex forms of stroke. Many individuals — in the range of several thousand — must be included in these studies to reach sufficient statistical power to reliably detect genetic associations. The latest and largest genome-wide association study (GWAS) of stroke, derived from the MEGASTROKE initiative[72], included 72,147 patients with stroke and 823,869 healthy controls[73].

To date, GWAS and a few large, robust, candidate gene association studies have identified 42 loci that are robustly associated with stroke — either all stroke or specific types or subtypes — at a genome-wide significant level[72–91] (TABLE 3). All associations were observed in populations of European ancestry or in meta-analyses that included multi-ancestry, but predominantly European ancestry, populations. Two loci that were identified as being genome-wide significant in the first published GWAS could not be confirmed in subsequent larger studies[75,84] (TABLE 3).

The loci identified include 22 associated with any stroke, 21 associated with any ischaemic stroke and 2 associated with ICH (TABLE 3). Considerable overlap is seen between the genome-wide significant loci associated with any stroke and those associated with any ischaemic stroke. Interestingly, one locus is associated with any ischaemic stroke and with deep ICH[72,76]; ICH associated with this locus is associated with underlying cerebral small-vessel disease, a leading cause of both ischaemic and haemorrhagic stroke[92]. In addition, 11 loci have been associated with ischaemic stroke subtypes — 7 with large artery stroke, 4 with cardioembolic stroke and 1 with small-vessel stroke[72]. Two identified loci were specifically associated with ischaemic stroke in young adults[86] and with cervical artery dissection[79], one of the leading aetiologies of stroke in young people (TABLE 3). Besides the genome-wide significant loci identified, additional robust associations at a lower significance threshold could provide further insight into

Table 3 | **Genetic loci associated with stroke**

SNP	Chromosome	Affected genes	Location	Risk allele (frequency)	Associated stroke subtype	Odds ratio (*P* value)
Risk of ischaemic stroke						
rs880315	1	*CASZ1*	Intronic	C (40%)	AS, AIS	1.05 (3.62×10^{-10})
rs12037987	1	*WNT2B*	Intronic	C (16%)	AS	1.07 (2.73×10^{-8})
rs146390073	1	*RGS7*	Intronic	T (2%)	CES	1.95 (2.20×10^{-8})
rs12476527	2	*KCNK3*	5′ UTR	G (48%)	AS	1.05 (6.44×10^{-8})[a]
rs7610618	3	*TM4SF4*, *TM4SF1*	Intergenic	T (1%)	LAS	2.33 (1.44×10^{-8})
rs34311906	4	*ANK2*	Intergenic	C (41%)	AIS	1.07 (1.07×10^{-8})
rs17612742	4	*EDNRA*	Intronic	C (21%)	LAS	1.19 (1.46×10^{-11})
rs6825454	4	*FGA*	Intergenic	C (31%)	AIS, AS	1.06 (7.43×10^{-10})
rs11957829	5	*LOC100505841*	Intronic	A (82%)	AIS, AS	1.07 (7.51×10^{-9})
rs6891174	5	*NKX2-5*	Intergenic	A (35%)	CES	1.11 (5.82×10^{-9})
rs16896398	6	*SLC22A7*,*ZNF318*	Intergenic	T (34%)	AS	1.05 (1.30×10^{-8})
rs42039	7	*CDK6*	3′ UTR	C (77%)	AIS, AS	1.07 (6.55×10^{-9})
rs7859727	9	Chr9p21	ncRNA_intronic	T (53%)	AS, AIS	1.05 (4.22×10^{-10})
rs10820405	9	*LINC01492*	ncRNA_intronic	G (82%)	LAS	1.20 (4.51×10^{-8})
rs2295786	10	*SH3PXD2A*	Intergenic	A (60%)	AS	1.05 (1.80×10^{-10})
rs7304841	12	*PDE3A*	Intronic	A (59%)	AIS	1.05 (4.93×10^{-8})
rs35436	12	*TBX3*	Intergenic	C (62%)	AS, AIS	1.05 (2.87×10^{-8})
rs9526212	13	*LRCH1*	Intronic	G (76%)	AS, AIS	1.06 (5.03×10^{-10})
rs4932370	15	*FURIN-FES*	Intergenic	A (33%)	AIS	1.05 (2.88×10^{-8})
rs11867415	17	*PRPF8*	Intronic	G (18%)	AIS	1.09 (4.81×10^{-8})
rs2229383	19	*ILF3*, *SLC44A2*	Exonic; synonymous substitution	T (65%)	AIS	1.05 (4.72×10^{-8})
rs8103309	19	*SMARCA4*, *LDLR*	Intergenic	T (65%)	AS	1.05 (3.40×10^{-8})
rs12124533	1	*TSPAN2*	Intergenic	T (24%)	LAS	1.17 (1.22×10^{-8})
rs1052053	1	*PMF1*, *SEMA4A*	Exonic; non-synonymous substitution	G (40%)	AS, AIS	1.06 (2.70×10^{-14})
rs13143308	4	*PITX2*	Intergenic	T (28%)	CES, AS, AIS	1.32 (1.86×10^{-47})
rs4959130	6	*FOXF2*	Intergenic	A (14%)	AS, AIS	1.08 (1.42×10^{-9})
rs2107595	7	*HDAC9*, *TWIST1*	Intergenic	A (24%)	LAS, AS, AIS	1.21 (3.65×10^{-15})
rs635634	9	*ABO*	Intergenic	T (19%)	AIS	1.08 (9.18×10^{-9})
rs2005108	11	*MMP12*	Intergenic	T (12%)	AIS, LAS	1.08 (3.33×10^{-8})
rs3184504	12	*SH2B3*	Exonic; non-synonymous	T (45%)	AIS, AS	1.08 (2.17×10^{-14})
rs12932445	16	*ZFHX3*	Intronic	C (21%)	CES	1.20 (6.86×10^{-18})
rs12445022	16	*ZCCHC14*	Intergenic	A (31%)	AS, AIS, SVS	1.06 (1.05×10^{-10})
rs1799983	7	*NOS3*	Exonic	T (32%)	AS	1.05 (2.2×10^{-8})
rs9521634	13	*COL4A1*	Intronic	C (36%)	AS	1.04 (3.8×10^{-8})
rs720470	21	*DYRK1A*	Intergenic	T (71%)	AS	1.05 (6.1×10^{-9})
rs112735431	17	*RNF213*	Exonic; non-synonymous	A (2%)	LAS (Japanese), AIS	3.58 (2.0×10^{-13})
rs4471613	15	*AQP9*	3′ UTR	A (2%)	AS (African-American)	0.82[b] (3.9×10^{-8})
rs11833579[c]	12	*NINJ2*	5′ UTR	A (23%)	AIS (incident stroke), AS	1.41 (2.3×10^{-10})
rs2230500[c]	14	*PRKCH*	Intronic	A (19%)	SVS (Japanese)	1.40 (5.1×10^{-7})[d]
rs11196288	10	*HABP2*	Intergenic	G (7%[e])	Young stroke	1.41 (9.5×10^{-9})
rs9349379	6	*PHACTR1*	Intronic	G (40%)	CeAD	0.77 (1.00×10^{-11})

Table 3 (cont.) | **Genetic loci associated with stroke**

SNP	Chromosome	Affected genes	Location	Risk allele (frequency)	Associated stroke subtype	Odds ratio (*P* value)
Risk of intracerebral haemorrhage						
rs2984613	1	*PMF1, SLC25A44*	Intronic	C (68%)	Deep ICH	1.33 (2.2×10^{-10})
rs429358	19	*APOE*	Exonic	ε2 (7%)	Lobar ICH	1.82 (6.6×10^{-10})
rs429358	19	*APOE*	Exonic	ε4 (12%)	Lobar and deep ICH[e]	2.20 (2.4×10^{-11})
Stroke severity and outcome						
rs1842681[f]	18	*LOC105372028*	Intronic	A (23%)	AIS	1.40 (5.27×10^{-9})
rs76221407	1	*PATJ*	Intronic	G (3%)	AIS	0.4[g] (1.72×10^{-9})
rs11655160	17	*PIRT*	Intergenic	G (76%)	ICH	0.82[h] (2.5×10^{-9})

Only associations that reached genome-wide significance ($P<5\times10^{-8}$ or log Bayes factor >6) are shown. Where the association was genome-wide significant for several stroke types or subtypes, the most significant *P* value is shown and the corresponding stroke (sub)type is listed first in the Associated stroke subtype column. All associations were observed in populations of European ancestry or in meta-analyses of transethnic populations that included predominantly those of European ancestry, unless specified otherwise. AS, all stroke; AIS, all ischaemic stroke; CeAD, cervical artery dissection; CES, cardioembolic stroke; ICH, intracerebral haemorrhage; LAS, large artery stroke; SNP, single-nucleotide polymorphism; SVS, small-vessel stroke; UTR, untranslated region. [a]log Bayesian factor >6 (6.47). [b]β = 0.82 indicating association with increased AS risk. [c]Could not be replicated in subsequent studies. [d]Considered genome-wide significant in the original publication that used a less dense genotyping array[75]. [e]Information from TOPMED and GnomAD. [f]A *trans*-expression quantitative trait locus for PPP1R21. [g]β = 0.4, indicating that an increase of 0.4 points in the modified Rankin scale score (reflecting a worse outcome) is attributed to each copy of the G allele. [h]β = 0.82, indicating an association with increased ICH volume; this SNP was also significantly associated with clinical severity (Glasgow Coma Scale score at admission) and functional outcome (modified Rankin scale score at 3 months).

the genetic contribution to stroke[72,93,94]. The existence of these associations with lower significance provides interesting evidence for a continuum between monogenic and multifactorial stroke; common variants in the same genes associated with Mendelian forms of stroke (for example, *COL4A1/COL4A2* and *HTRA1*) are also associated with complex forms of ischaemic stroke and ICH[72,94].

Genetic studies of stroke have included either large numbers of participants with any type of stroke or smaller numbers of participants with highly specific subtypes of stroke; both approaches have proven successful and have produced complementary findings. Genes associated with any stroke could represent either a very strong association with a specific, common subtype of stroke (for example, the *HDAC9* and *PITX2* loci are very strongly associated with large artery stroke and cardioembolic stroke, respectively) or might represent involvement of biological pathways that influence the risk of all or most types of stroke. These pathways could involve: genes that predispose to vascular risk factors for all stroke types, such as high blood pressure; genes that predispose to vessel fragility, thrombosis or bleeding; or genes that modulate the tolerance and resilience to brain injury[72] (FIG. 2).

The main goal of identifying genetic associations with stroke is to improve our understanding of the molecular pathways that underlie stroke and its subtypes and identify novel drug targets. To achieve this goal, statistical associations of specific loci need to be complemented by functional studies to identify the most compelling candidates[72]. Identifying genetic variations that are shared between stroke and other complex phenotypes can also tell us more about disease mechanisms[72,79,95]. The ability to improve prediction of risk on the basis of genetic associations has long been considered an unrealistic goal for complex conditions, such as stroke, owing to the small effect size of individual variants. However, development of polygenic risk scores has demonstrated that these variants can be used to identify individuals with a risk equivalent to that in monogenic conditions[96]. This approach has shown that 8% of the general population has a risk of coronary artery disease that is threefold higher than average[96]. This approach remains to be applied to stroke.

Severity, outcome and aetiology. Two large, collaborative meta-analyses of GWAS from the Genetics of Ischaemic Stroke Functional Outcome (GISCOME) network and the ISGC have identified common and low-frequency genetic variants at two distinct loci that are associated with functional outcome (modified Rankin scale score) 3 months after stroke[97]. An intergenic region on chromosome 17p12, which seems to be greatly affected by copy number variation, has also been associated with haematoma volume, clinical severity and functional outcome in non-lobar ICH[98]. Genetic imbalance — the result of protein-coding genes being affected by copy number variations — has also been associated with unfavourable functional outcomes at 3 months after an ischaemic stroke[99].

Overall, the genetic architectures of stroke risk and early neurological changes after stroke are distinct[100]. However, some loci that are associated with stroke risk might also influence stroke severity if they modulate the extent of the underlying pathology, particularly cerebral small-vessel disease, which can cause stroke but also affects stroke severity and outcome[91,101]. For example, the *APOE* locus, which is associated with small-vessel disease, has also been associated with mean volume of an ICH — carriers of the *APOE*ε2* allele had larger ICH volumes[85]. Similarly, the *FOXF2* locus, which is involved

in cerebral vessel development and associated with an increased risk of any stroke, has also been associated with a higher risk of death from stroke[91].

Little research has been done on genetic associations with different stroke aetiologies, but one study has shown that, when combined in a polygenic risk score, genetic risk variants associated with atrial fibrillation were associated with cardioembolic stroke and stroke of undetermined aetiology[102]. Further studies are needed to determine whether genetic risk variants for atrial fibrillation can serve as diagnostic biomarkers for strokes caused by atrial fibrillation.

Stroke epigenomics. In contrast with genetic factors, which are fixed throughout the lifetime, epigenetic modifications — the most common of which is DNA methylation — that regulate gene expression are dynamic and tissue-specific. Epigenome-wide association studies enable powerful assessment of how epigenetic factors influence complex diseases[103,104]. DNA methylation could influence mechanisms of stroke by mediating genetic or environmental risk factors and/or could serve as biomarkers of stroke if epigenetic changes are a cause or consequence of the disease (both directions of association are possible). Epigenome-wide association studies of stroke and related phenotypes are still in their infancy but findings to date are promising. For example, differential methylation patterns in *PPM1A*, which encodes protein phosphatase 1A, has been associated with vascular recurrence in aspirin-treated stroke patients[105]. However, no global differences in methylation have been found between different ischaemic stroke subtypes[106].

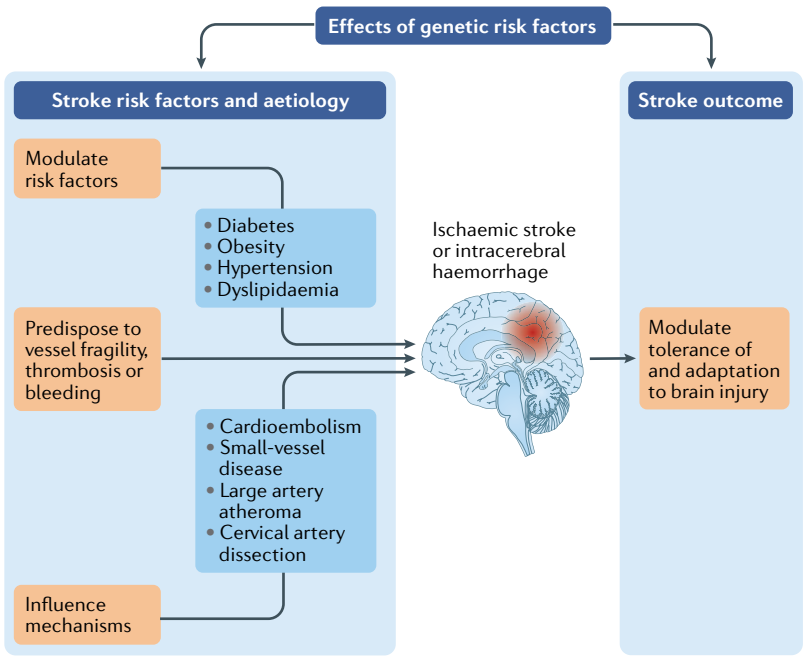

Fig. 2 | **Mechanisms by which genetic variants can influence the risk of stroke and stroke outcomes.** Before stroke (left), genetic factors influence the development of risk factors (such as diabetes, obesity, hypertension and dyslipidaemia), an individual's vascular vulnerability and the mechanism by which stroke can develop. After stroke (right), genetic factors influence how much damage occurs as a result of the insult and how the brain adapts and recovers.

DNA methylation can also be used to estimate biological age, which is an independent predictor of ischaemic stroke outcome regardless of chronological age[107].

Future directions of stroke genomics. Most published studies of complex stroke genetics have focused on common or low-frequency single-nucleotide polymorphisms, which explain only a small proportion of stroke heritability. Other types of variation, such as rare (present in less than <1% of the population) single-nucleotide variants or structural variations (for example, copy number variants), have been insufficiently explored[108]. Several efforts to address this shortfall through use of next-generation sequencing, such as the Trans-Omics for Precision Medicine (TOPMed) programme[109] and other initiatives, are currently underway.

Genetic studies should also be expanded to include groups that are not of European ancestry. This step is essential to enhance the discovery of stroke risk loci, to identify causal variants and genes and to explore how representative the identified associations are of the wider population[110]. Finally, the occurrence of stroke and its severity are consequences of complex interplay between factors beyond the DNA sequence, including epigenetic modifications, RNA transcripts, proteins and metabolites[111]. Therefore, our understanding of the biology that underlies the risk of stroke and stroke outcomes will be enriched by combining genomics information with epigenomics, transcriptomics, metabolomics and proteomics data.

Stroke transcriptomics

Transcriptomics studies focus on gene expression in peripheral blood cells and on RNAs in the plasma as potential biomarkers. An advantage of RNA is that all known coding and non-coding transcripts can be assessed by microarray or RNA sequencing. The identified RNA can be confirmed using a second method, such as PCR with reverse transcription (RT-PCR). This approach has led to the identification of several transcript panels that are associated with stroke diagnosis, stroke aetiology and stroke risk (TABLE 4). However, studies of RNA expression in stroke remain scarce. Initial studies, as described below, provide proof of principle that differences in gene expression and in expression of microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) have potential as stroke biomarkers. They also provide novel insight into the biology of human stroke and could facilitate risk classification, diagnosis of stroke and determination of stroke aetiology. Further evaluation and validation of the identified markers in large cohorts is required for their translation into clinical practice.

RNA for stroke diagnosis. Patterns of leukocyte RNA expression have been identified that distinguish patients with ischaemic stroke from those with haemorrhagic stroke and stroke mimics and from healthy controls. Several panels of gene transcripts (9–97 genes) expressed in blood cells can distinguish patients with ischaemic stroke from those with haemorrhagic stroke and from controls[112,113], and enable identification of ischaemic

Table 4 | **Potential stroke biomarkers identified in transcriptomics studies**

RNA transcript	Potential use	Evidence	Refs
mRNA	Diagnosis of stroke	Multigene panels (9–97 genes) reflect aspects of the peripheral immune response and thrombus formation associated with ischaemic stroke	112–114
	Diagnosis of transient ischaemic attack (TIA)	A 34-gene panel enabled differentiation between patients with TIA and healthy controls, and a 26-gene panel enabled differentiation between TIA and mimics	171,172
	Identification of stroke aetiology	Gene panels distinguished between cardioembolic and large-vessel stroke (23-gene and 40-gene panels), stroke with atrial fibrillation and without atrial fibrillation (37-gene panel), and lacunar and non-lacunar stroke (41-gene panel), and determined the cause in cryptogenic stroke (40-gene and 41-gene panels)	115–118
Long non-coding RNA	Diagnosis of stroke	Differences in long non-coding RNA in blood cells between patients with stroke and healthy controls	128–130
	Identification of stroke aetiology	Differences in long non-coding RNA between healthy controls and patients with large-vessel atherosclerosis	131,132
Extracellular microRNA	Diagnosis of stroke	Varied results: decreased levels of miR-32-3p, miR-106b-5p, miR-423-5p, miR-451a, miR-1246, miR-1299, miR-3149 and miR-4739, and increased levels of miR-224-3p, miR-377-5p, miR-518b, miR-532-5p and miR-1913 associated with stroke	122
Intracellular microRNA	Diagnosis of stroke	Decreased levels of miR-122, miR-148a, let-7i, miR-19a, miR-320d and miR-4429, and increased levels of miR-363 and miR-487b associated with stroke	124

stroke with sensitivities and specificities of 80–90%. The gene transcripts included in these panels, which include those that encode S100B, arginase, factor V, CD28 and nuclear factor of activated T cells (NFAT), largely relate to differences in immune function and thrombosis[114].

Additional work is required to develop RNA into a diagnostic tool in the context of acute stroke. Methods to measure RNA in clinical practice already exist for several diseases (for example, HIV, breast cancer and cervical cancer), but the turnaround time from blood sample acquisition to result is 4–24 h. This time needs to be reduced for application to stroke management because any delay to reperfusion therapies increases brain injury and disability. The best tool for measurement of RNA in 10–20 min remains to be determined — RT-PCR has potential, but other strategies in development include functionalized nanowires, electrochemical methods and antibody–oligonucleotide conjugates.

RNA and stroke aetiology. Panels of RNA have been developed for the identification of likely cardioembolic, large-vessel or small-vessel causes of stroke[115–117]. One 40-gene panel enabled differentiation of cardioembolic stroke from large-vessel stroke with sensitivity and specificity of >95%[115]. Similarly, a 37-gene panel enabled differentiation of cardioembolic stroke due to atrial fibrillation from cardioembolic stroke without atrial fibrillation with sensitivity and specificity of 90%. In addition, a 41-gene panel enabled discrimination of lacunar from non-lacunar stroke with sensitivity and specificity of >90%[117]. The genes included in these panels indicate differences in inflammation and clot formation between stroke subtypes. Of interest, application of such RNA panels to patients with stroke of unclear aetiology indicated that 27% had stroke with atrial fibrillation and 18% had large-vessel stroke[118]. However, further evaluation of RNA panels in larger cohorts is required to refine them and determine their clinical implications. For example, it will be important to demonstrate that patients who are identified with RNA panels as having cardioembolic or large-vessel stroke benefit from therapies that target these aetiologies; such studies remain to be performed.

MicroRNA in stroke. Both extracellular plasma miRNAs and intracellular miRNAs have been studied for their potential as biomarkers in stroke. miRNAs are of interest in this context because they regulate cellular gene expression and can therefore have important signalling roles in stroke. Multiple studies have identified changes in extracellular miRNAs in stroke[119–122]. These plasma miRNAs derive from the ischaemic brain or from the heart, lungs, kidneys, leukocytes and/or other tissues. However, results have been inconsistent, possibly owing to differences in the methods used to isolate and measure RNA, and to platelet contamination[122,123].

The intracellular miRNAs that are altered after stroke include miR-122, miR-148a, let-7i, miR-19a, miR-320d, miR-4429, miR-363 and miR-487b[124]. To date, many of the intracellular miRNAs that have been identified as being altered after stroke are downregulated, an observation that is consistent with an increase in leukocyte mRNA expression after stroke, given the inhibitory effects of mRNA. These miRNAs are predicted to regulate Toll-like receptor signalling, nuclear factor-κB (NF-κB) signalling, leukocyte extravasation signalling and the prothrombin activation pathway[124]. This suggests that miRNA regulates the immune response to ischaemic brain injury after stroke, which may have implications for neuroprotection[119,120,122,125–127].

Long non-coding RNA in stroke. As their name implies, lncRNAs are RNA molecules that are longer than miRNAs (>200 bp) and do not code for proteins but

contribute to gene expression. As such, lncRNAs are likely to have important roles in stroke. In one study, differential expression in blood was seen for 299 lncRNAs in men with stroke and for 97 lncRNAs in women with stroke, in comparison with healthy controls[128]. Some of the lncRNAs identified were associated with stroke risk genes, including those that encode lipoprotein, lipoprotein(a)-like 2, prostaglandin I_2 synthase, α-adducins and the ABO blood group (transferase A, α1-3-*N*-acetylgalactosaminyltransferase; transferase B, α1-3-galactosyltransferase)[128]. Another study identified 70 upregulated and 128 downregulated lncRNAs in patients with ischaemic stroke compared with healthy controls[129]. The dysregulation of three of these lncRNAs (linc-DHFRL1-4, SNHG15 and linc-FAM98A-3) was confirmed by RT-PCR, and outperformed the ability of brain-derived neurotrophic factor (BDNF) and neuron-specific enolase (NSE) to identify stroke, with an area under the receiver operating characteristic curve of 0.84 (REF.[129]). Changes in lncRNAs have also been associated with Moyamoya disease[130] and large-vessel stroke[131,132].

Stroke metabolomics

Brain ischaemia causes local and systemic metabolic alterations, such as changes in cellular energy metabolism pathways and a global stress response. The discovery of these alterations has triggered studies to investigate whether metabolites can serve as circulating biomarkers for estimating stroke risk, defining stroke diagnosis and identifying aetiological stroke subtypes (TABLE 5). By contrast, the value of metabolites for prediction of functional outcomes after stroke has not been systematically studied.

Table 5 | **Potential metabolite biomarkers in stroke**

Metabolite	Class or origin of metabolite	Finding	Ref.
Estimation of stroke risk			
Linoleic acid	ω-6 Polyunsaturated fatty acid	Higher linoleic acid levels associated with lower risk of incident cardiovascular disease and incident ischaemic stroke	136
Tetradecanedioate Hexadecanedioate	ω-Oxidation derived long-chain dicarboxylic acids	Higher levels of tetradecanedioate and hexadecanedioate associated with incident cardioembolic stroke	138
Diagnosis of stroke			
Ceramide 42:1 Sphingomyelin 36:0	Sphingomyelins	Enabled differentiation between ischaemic stroke and stroke mimics in a small number of patients	140
Branched chain amino acids	Catabolism	Valine, leucine and isoleucine levels lower in patients with cardioembolic stroke than in patients with transient ischaemic attack	141
Identification of aetiology			
Total free fatty acids	Lipids	Plasma concentration of total free fatty acids higher in patients with cardioembolic stroke than in patients with non-cardioembolic stroke	142
Succinate	Tricarboxylic acid cycle	High circulating levels of succinate associated with cardioembolic stroke, atrial dysfunction and left atrial enlargement	145

Metabolites are small molecules (typically <1.5 kDa) from entities such as lipids, amino acids, carbohydrates and nucleotides. The number of known circulating metabolites is >25,000, but levels as low as ~3,000 have been quantified[133]. The composition of circulating metabolites is a downstream function of gene, expression and environmental changes, such as dietary intake and variation in gut microbiota[134]. Profiling of the circulating metabolome could, therefore, capture systems-level information. Furthermore, metabolites regulate key biological processes, such as cell signalling, so the metabolome could provide information about disease mechanisms.

To account for the fact that metabolites differ greatly in their polarity, charges and sizes, metabolomics profiling involves different technologies, most often based on NMR spectroscopy and mass spectrometry[135]. NMR spectroscopy requires minimal sample preparation and is highly reproducible, but can only detect the most abundant metabolites in a sample. Mass spectrometry-based methods are more sensitive and can detect thousands of metabolites, but are more susceptible to variability, more expensive and require more complex bioinformatics analyses. Use of internal standards and approaches that are tailored to the metabolites of interest (a targeted approach) enables absolute levels of specific metabolites to be determined[135]. By contrast, an absence of internal standards and use of broad analytical approaches that capture metabolites with different polarities, sizes and charges (an untargeted approach) enable relative quantification of as many metabolites in the sample as possible, including unknown compounds.

Metabolites and risk of stroke. Established risk factors for stroke, such as dyslipidaemia and diabetes mellitus, are characterized by metabolic changes. Therefore, direct quantification of metabolite levels might be a more accurate way to estimate stroke risk than traditional classifications based on risk factors. Studies of metabolites in this way could also provide insight into preventive strategies, such as medication or dietary intake. For example, in a study published in 2019, data from 30 prospective observational studies (including 68,659 patients) were pooled to study the association between levels of linoleic acid — the main dietary ω-6 polyunsaturated fatty acid — and cardiovascular disease[136]. Higher circulating levels of linoleic acid were associated with a lower risk of all cardiovascular disease and of ischaemic stroke alone. This finding can be used to inform dietary recommendations.

In an earlier study, 135 lipids were profiled in 685 participants in the prospective population-based Bruneck study, and the findings were validated in the TwinsUK cohort (1,453 individuals)[137]. Of the lipids profiled, high circulating levels of triacylglycerols and cholesterol esters with low carbon numbers and double-bond content were the strongest predictors of a composite cardiovascular end point that included stroke. In the same study, addition of three of the associated lipids

to a predictive model based on traditional risk factors improved risk discrimination and 10-year risk reclassification. These findings highlight the need for lipid management to address lipid composition, in contrast to current practice, which focuses on cholesterol and total lipid levels.

In another study published in 2019, the metabolic profiles of 3,904 individuals from the Atherosclerosis Risk in Communities (ARIC) study were studied with incident ischaemic stroke as a sole end point[138]. Among the 384 metabolites quantified, higher levels of the long-chain dicarboxylic acids tetradecanedioate and hexadecanedioate were associated with a higher risk of incident ischaemic stroke, and specifically with cardioembolic stroke. Both metabolites are products of the ω-oxidation of fatty acids, but how they are functionally linked to cardioembolic stroke remains unclear.

Metabolites for stroke diagnosis. Metabolites are thought to be able to cross the BBB more easily than proteins owing to their smaller size. In addition, the wide range of half-lives among metabolites means that their circulating levels could reflect acute-onset diseases earlier than other molecular entities can. On this basis, multiple studies have been conducted to determine whether metabolites could be diagnostic biomarkers in the acute phase of ischaemic stroke. However, only three studies have addressed this question in a clinically relevant setting — that is, within the first 12 h after symptom onset and with clinically relevant control groups, such as patients with stroke mimics, transient ischaemic attack or ICH rather than healthy controls.

In the most recent of these three studies, blood samples from patients with ischaemic stroke or ICH were analysed with mass spectrometry. A set of five variates — levels of asparagine and tiglylcarnitine, and three metabolite ratios — correctly identified 79% of patients with stroke[139]. The study included >200 patients and the findings were internally validated, but metabolite levels were not adjusted for demographic and vascular risk factors and no external validation was done, which limits the generalization of these findings.

By contrast, a targeted approach was used in an earlier study[140] to focus on sphingolipids, which are highly abundant in the brain and readily cross the BBB. On the basis of data from rodent models of experimental stroke and traumatic brain injury, the specific lipids selected for analysis were ceramide 42:1 and sphingomyelin 36:0. Analysis of these markers enabled differentiation of patients with ischaemic stroke from those with stroke mimics with an area under the curve of 0.8712. However, only 14 patients were included, thus limiting the strength of the final conclusions.

In the other of these three studies, mass spectrometry analysis of 68 metabolites in plasma after experimental stroke showed a reduction in levels of branched-chain amino acids that depended on infarct size, and this set of amino acids was also reduced in patients with cardioembolic stroke compared with patients with transient ischaemic attack after adjusting for demographic factors[141]. Future analyses will determine how well this metabolite set identifies all ischaemic stroke subtypes among patients with stroke-like symptoms upon arrival at hospital.

Metabolites and stroke aetiology. Some studies have been conducted to examine whether metabolites can be used to differentiate between aetiological stroke subtypes. In one of these studies, plasma levels of total free fatty acids were quantified in 669 patients with acute ischaemic stroke[142]. Levels of free fatty acids were higher in patients with cardioembolic stroke than in patients with other types of ischaemic stroke (according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification)[143] after adjustment for various confounders, such as the NIHSS score at baseline and vascular risk factors. Whether this observation would be sufficient to guide clinical decision-making regarding secondary prophylactic treatment remains unclear.

In another study, 144 circulating metabolites were quantified with mass spectrometry in samples from 367 patients with acute ischaemic stroke and individual metabolites were assessed for their ability to distinguish between cardioembolic and non-cardioembolic strokes (according to the causative classification of stroke)[144]. Levels of the tricarboxylic acid cycle intermediates succinate, α-ketoglutarate and malate were higher in patients with cardioembolic stroke than in patients with non-cardioembolic stroke. Succinate was also associated with left atrial enlargement and subclinical atrial dysfunction[145], indicating that energy metabolism is altered in patients with cardioembolism and structural cardiac abnormalities, thereby supporting the abnormal atrial substrate model of cardioembolic stroke[146].

Inconsistencies in metabolomics studies. Collectively, several circulating metabolites have been convincingly associated with stroke risk and diagnosis. However, inconsistencies exist between the associated metabolites, which might be the result of several study design features. First, small sample sizes results in insufficient statistical power to capture true differences between hundreds of measured metabolites. Second, different recruitment strategies in different studies result in differences between the groups of patients included. Third, the timing of blood sampling varies between and within studies. Last, different detection platforms, such as targeted and untargeted approaches, have been used in different studies. Consideration of these methodological confounders in future studies, together with technological advances[147], will move metabolomics closer to clinical applications.

Integromics and systems biology

Combining omics data. The progress of all omics technologies has necessitated development of tools to facilitate analysis and interpretation of the multidimensional data being generated. Many statistical methods have been developed for independent analysis of large-scale, high-quality data from each level of omics, but such individual interpretations overlook the crosstalk between different molecular entities and could miss biologically relevant information. Consequently, integrated analysis of data obtained with different omics approaches — here

referred to as integromics — is becoming crucial for a deep understanding of pathological processes in a biologically meaningful context.

Integromics is expected to provide new insights into complex biological systems and to reveal interaction networks between processes at all molecular levels. Integromics-based biomarker candidates would be biologically relevant regardless of whether the changes at each single omics level are large or small. Biomarker candidates for which the changes are extremely large are most likely to be used as stroke biomarkers in clinical practice because the large change would make it easier to establish a cut-off point and discriminate between conditions. Candidate biomarkers for which changes are large or moderate could also be considered as therapeutic targets.

Little is known about integromics approaches in the context of ischaemic stroke specifically, but multivariate statistical methods for integromics applied to other diseases are continuously emerging. Most of these methods include data normalization and dimension-reduction approaches, such as principal component analysis, which break down data into a few variables to facilitate identification of those that best explain phenotypic differences between patients with stroke. Other multivariate analysis methods — including canonical correlation analysis and all subsequently described versions — have been designed to investigate overall correlations between sets of variables and to ultimately identify the factors that best account for a specific biological condition. Other integromics frameworks include partial least squares regression analysis[148] or multiple factor analysis[149], which enable identification of the main sources of phenotypic variability between conditions. In all cases, however, difficulties can arise in discerning between biologically relevant and irrelevant molecules because the combined effects of multiple factors and high variability in individual datasets can lead to artefacts[149].

Systems biology. Systems biology is emerging as a more sophisticated approach to integromics. This approach combines experimental data at multiple molecular levels with computational modelling and treats the system as a whole to facilitate identification of data with promising diagnostic, prognostic or therapeutic value[150]. Systems biology is also designed to track specific molecular interactions over time rather than provide a static map of molecular relationships. Given the complexity and heterogeneity of stroke, understanding how these molecular interactions evolve over time could help to determine the optimal time point for a biomarker measurement or the therapeutic time window of action of a particular drug.

Increasingly, bioinformatics resources are being used to integrate molecular data with clinical information (that is, non-omics data). This aspect is important because external biological and clinical knowledge can help to guide interpretation of molecular data by providing a global view of the disease. However, combining such information in the context of stroke is complicated in practice and approaches are not yet mature. Clinical information, including diagnosis, aetiology and prognosis, is easily accessible, but in the context of human stroke, omics techniques have mainly been applied to brain samples from people who have died. Molecular signatures in the brain after death might differ substantially from those in the early phases of the disease, which would be more clinically relevant.

In addition, confounders are a concern that needs to be addressed at the time of integrating omics data with clinical and demographic information about patients with stroke. With the advent of large GWAS and the emergence of new candidate circulating biomarkers, the use of (two-sample) Mendelian randomization to provide evidence for or rule out causal associations between these circulating biomarkers and stroke has become increasingly popular[151–157]. This technique can be useful, but caution is needed in its application and interpretation. For example, bidirectional Mendelian randomization can be used to distinguish biomarkers that are most likely causal and those that are most likely a consequence of stroke[158]. For example, use of this technique has indicated that higher circulating levels of lipids[157], homocysteine[153] and CC-chemokine ligand 2 (CCL2; also known as MCP1)[155,156] are associated with an increased risk of stroke. Mendelian randomization can also be used to estimate the effects of an existing drug on a given stroke risk factor by investigating how genetic variants that modulate expression of the drug target are associated with the disease. A similar approach can be used to predict adverse effects[159]. However, several potential biases need to be considered in the application of Mendelian randomization[160,161]. One prominent example is horizontal pleiotropy, in which a variant influences the disease outcome in ways unrelated to its effect on the exposure, which can distort estimates of causal relationships, although increasingly elaborate methods are being developed to negate this effect[162]. Consequently, evidence beyond such analyses, including experimental evidence, is required to prove that a given circulating biomarker represents a therapeutic target. Work to obtain such evidence for CCL2 is underway[156].

Overall, although integrative approaches are clearly improving knowledge, approaches to combine omics data with phenotypic data in stroke need to be developed further. Once developed and implemented, however, these approaches will enable prioritization of specific pathways and gene targets in the design of optimal diagnostic, prognostic and therapeutic strategies.

Stroke treatment. Given that stroke is a complex, multifactorial disease in which many biological processes are simultaneously deregulated, modulation of a single molecular factor is unlikely to be sufficient to attenuate or reverse the progression of stroke pathology. For this reason, research efforts are focusing on identification of key molecular signatures that include sets of important contributors to stroke pathology, a process similar to the development of combinational therapies that are now in use for cancer[163,164]. In this context, integromics is expected to facilitate identification of biologically interconnected processes in stroke that can be simultaneously modulated with combinational therapies.

Integromics could also become important in the development of personalized therapy. Alignment of clinical phenotypes with underlying, multilevel molecular networks could facilitate comparisons of the biological signatures of clinical manifestations. Molecular subtyping with multi-omics approaches is already used in cancer to identify subtypes, make prognoses and identify disease mechanisms to identify and administer effective personalized treatments[165,166]. Multi-omics and integration with clinical data are also being investigated as a way to accelerate precision medicine and personalized healthcare in autism spectrum disorders[167].

In addition, we have begun to see examples of how multi-omics can be used to build in silico models for prediction of changes that take place in some diseases. For example, in one study[168], the effect of stroke was simulated by decreasing uptake levels of metabolites into the brain, which created changes in glutamate levels in the model that were similar to experimental observations in reversible middle cerebral artery occlusion. Such multi-omics-based models will enable comparison of biomarkers and pathways — the so-called diseasome — that are shared between diseases, leading to drug repurposing opportunities.

Conclusions

Technological advances have led to the 'omics era', which is enabling the collection and integration of data and information at different molecular levels. The information obtained through omics techniques will contribute to a better understanding of stroke pathophysiology, will offer new opportunities for diagnosis and prognosis and will lead to improved management of patients with stroke. All the emerging omics approaches discussed above will undoubtedly transform medical practice in the near future.

Further research must be done to validate findings before integrating them into clinical practice. This validation must include more diverse study populations, as a lack of ethnic and geographical diversity among participants in published studies is currently an important bias. This shortcoming is being addressed in some stroke genetics studies, such as those conducted by the Consortium of Minority Population Genome-Wide Association Studies of Stroke (COMPASS) and the South London Ethnicity and Stroke Study (SLESS)[88,169]. In stroke proteomics, diversity is not being addressed routinely despite the fact that results from the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study indicate that associations between some proteins and the risk of stroke differ depending on ethnicity[170].

Although confirmatory experiments and validations are required, there is a clear path to clinical application of multi-omics approaches in the management of stroke. If the unmet needs in the field are addressed, these approaches will provide opportunities to implement biomarkers at several stages of the stroke care pathway, with the potential to transform stroke management and dramatically improve patient outcomes.

Published online 22 April 2020

1. Benjamin, E. J. et al. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. *Circulation* **139**, e56–e528 (2019).
2. Ramiro, L., Simats, A., García-Berrocoso, T. & Montaner, J. Inflammatory molecules might become both biomarkers and therapeutic targets for stroke management. *Ther. Adv. Neurol. Disord.* **11**, 1–24 (2018).
3. Kunz, A. et al. Effects of ultraearly intravenous thrombolysis on outcomes in ischemic stroke: the STEMO (Stroke Emergency Mobile) group. *Circulation* **135**, 1765–1767 (2017).
4. Senn, R., Elkind, M. S. V., Montaner, J., Christ-Crain, M. & Katan, M. Potential role of blood biomarkers in the management of nontraumatic intracerebral hemorrhage. *Cerebrovasc. Dis.* **38**, 6 (2014).
5. Dirnagl, U., Iadecola, C. & Moskowitz, M. A. Pathobiology of ischaemic stroke: an integrated view. *Trends Neurosci.* **22**, 391–397 (1999).
6. Wetterling, F. et al. Investigating potentially salvageable penumbra tissue in an in vivo model of transient ischemic stroke using sodium, diffusion, and perfusion magnetic resonance imaging. *BMC Neurosci.* **17**, 1–10 (2016).
7. Chamorro, Ã. et al. The immunology of acute stroke. *Nat. Rev. Neurol.* **8**, 401–410 (2012).
8. Iadecola, C. & Anrather, J. The immunology of stroke: from mechanisms to translation. *Nat. Med.* **17**, 796–808 (2011).
9. Simats, A., García-Berrocoso, T. & Montaner, J. Neuroinflammatory biomarkers: from stroke diagnosis and prognosis to therapy. *Biochim. Biophys. Acta* **1862**, 411–424 (2016).
10. Yan, Y.-Y. et al. Immune cells after ischemic stroke onset: roles, migration, and target intervention. *J. Mol. Neurosci.* **66**, 342–355 (2018).
11. Sas, A., Horvath, L., Olah, C. & Valikovics, A. in *Mechanisms of Neuroinflammation* (ed. Aranda Abreu, G. E.) 119–144 (IntechOpen, 2017).
12. Perez-de-Puig, I. et al. Neutrophil recruitment to the brain in mouse and human ischemic stroke. *Acta Neuropathol.* **129**, 239–257 (2015).
13. Lakhan, S. E., Kirchgessner, A., Tepper, D. & Leonard, A. Matrix metalloproteinases and blood–brain barrier disruption in acute ischemic stroke. *Front. Neurol.* **4**, 32 (2013).
14. Lambertsen, K. L., Finsen, B. & Clausen, B. H. Post-stroke inflammation — target or tool for therapy? *Acta Neuropathol.* **137**, 693–714 (2018).
15. Pennypacker, K. R. Targeting the peripheral inflammatory response to stroke: role of the spleen. *Transl Stroke Res.* **5**, 635–637 (2015).
16. Liu, Q. et al. Brain ischemia suppresses immunity in the periphery and brain via different neurogenic innervations. *Immunity* **46**, 474–487 (2017).
17. Chavan, S. S., Pavlov, V. A. & Tracey, K. J. Mechanisms and therapeutic relevance of neuro-immune communication. *Immunity* **46**, 927–942 (2017).
18. Dirnagl, U. et al. Stroke-induced immunodepression: experimental evidence and clinical relevance. *Stroke* **38**, 770–773 (2007).
19. Esmaeili, A., Dadkhahfar, S., Fadakar, K. & Rezaei, N. Post-stroke immunodeficiency: effects of sensitization and tolerization to brain antigens. *Int. Rev. Immunol.* **31**, 396–409 (2012).
20. Shi, K., Wood, K., Shi, F. D., Wang, X. & Liu, Q. Stroke-induced immunosuppression and poststroke infection. *Stroke Vasc. Neurol.* **3**, 34–41 (2018).
21. Westendorp, W. F., Nederkoorn, P. J., Vermeij, J.-D., Dijkgraaf, M. G. & van de Beek, D. Post-stroke infection: a systematic review and meta-analysis. *BMC Neurol.* **11**, 110 (2011).
22. Montaner, J. et al. A panel of biomarkers including caspase-3 and D-dimer may differentiate acute stroke from stroke-mimicking conditions in the emergency department. *J. Intern. Med.* **270**, 166–174 (2011).
23. Laskowitz, D. T., Kasner, S. E., Saver, J., Remmel, K. S. & Jauch, E. C. Clinical usefulness of a biomarker-based diagnostic test for acute stroke: the Biomarker Rapid Assessment in Ischemic Injury (BRAIN) study. *Stroke* **40**, 77–85 (2009).
24. Bustamante, A. et al. Blood biomarkers for the early diagnosis of stroke: the Stroke-Chip study. *Stroke* **48**, 2419–2425 (2017).
25. Penn, A. M. et al. Validation of a proteomic biomarker panel to diagnose minor-stroke and transient ischaemic attack: phase 2 of SpecTRA, a large scale translational study. *Biomarkers* **23**, 793–803 (2018).
26. Ebinger, M. et al. Effect of the use of ambulance-based thrombolysis on time to thrombolysis in acute ischemic stroke: a randomized clinical trial. *JAMA* **311**, 1622–1631 (2014).
27. Eng, L. F., Ghirnikar, R. S. & Lee, Y. L. Glial fibrillary acidic protein: GFAP-thirty-one years (1969–2000). *Neurochem. Res.* **25**, 1439–1451 (2000).
28. Perry, L. A. et al. Glial fibrillary acidic protein for the early diagnosis of intracerebral hemorrhage: systematic review and meta-analysis of diagnostic test accuracy. *Int. J. Stroke* **14**, 390–399 (2018).
This paper presents a meta-analysis of the accuracy of GFAP, one of the most promising candidates for diagnosis of acute ICH.
29. Llombart, V. et al. Plasmatic retinol-binding protein 4 and glial fibrillary acidic protein as biomarkers to differentiate ischemic stroke and intracerebral hemorrhage. *J. Neurochem.* **136**, 416–424 (2016).
30. Montaner, J. et al. Matrix metalloproteinase expression is related to hemorrhagic transformation after cardioembolic stroke. *Stroke* **32**, 2762–2767 (2001).
31. Castellanos, M. et al. Plasma metalloproteinase-9 concentration predicts hemorrhagic transformation in acute ischemic stroke. *Stroke* **34**, 40–45 (2003).
32. Yuan, R. et al. Predictive value of plasma matrix metalloproteinase-9 concentrations for spontaneous haemorrhagic transformation in patients with acute ischaemic stroke: a cohort study in Chinese patients. *J. Clin. Neurosci.* **58**, 108–112 (2018).
33. Wang, L. et al. The accuracy of serum matrix metalloproteinase-9 for predicting hemorrhagic transformation after acute ischemic stroke: a systematic review and meta-analysis. *J. Stroke Cerebrovasc. Dis.* **27**, 1653–1665 (2018).

34. Castellanos, M. et al. Plasma cellular-fibronectin concentration predicts hemorrhagic transformation after thrombolytic therapy in acute ischemic stroke. *Stroke* **35**, 1671–1676 (2004).
35. Hernandez-Guillamon, M. et al. Plasma VAP-1/SSAO activity predicts intracranial hemorrhages and adverse neurological outcome after tissue plasminogen activator treatment in stroke. *Stroke* **41**, 1528–1535 (2010).
36. Ribo, M. et al. Admission fibrinolytic profile is associated with symptomatic hemorrhagic transformation in stroke patients treated with tissue plasminogen activator. *Stroke* **35**, 2123–2127 (2004).
37. Bustamante, A. et al. Usefulness of ADAMTS13 to predict response to recanalization therapies in acute ischemic stroke. *Neurology* **90**, e995–e1004 (2018).
This pilot study considers the usefulness of blood biomarkers to guide both intravenous and endovascular reperfusion therapies.
38. De Marchis, G. M. et al. Copeptin adds prognostic information after ischemic stroke: results from the CoRisk study. *Neurology* **80**, 1278–1286 (2013).
39. Bustamante, A. et al. Ischemic stroke outcome: a review of the influence of post-stroke complications within the different scenarios of stroke care. *Eur. J. Intern. Med.* **29**, 9–21 (2016).
40. Hoffmann, S. et al. Stroke-induced immunodepression and dysphagia independently predict stroke-associated pneumonia — the PREDICT study. *J. Cereb. Blood Flow Metab.* **37**, 3671–3682 (2017).
41. Wang, C. et al. Procalcitonin is a stronger predictor of long-term functional outcome and mortality than high-sensitivity C-reactive protein in patients with ischemic stroke. *Mol. Neurobiol.* **53**, 1509–1517 (2016).
42. Ulm, L. et al. The randomized controlled STRAWINSKI trial: procalcitonin-guided antibiotic therapy after stroke. *Front. Neurol.* **8**, 153 (2017).
This clinical trial with a remarkable design might represent an example of how to implement biomarkers in clinical practice, despite negative findings.
43. Azurmendi, L. et al. Proteomic discovery and verification of serum amyloid A as a predictor marker of patients at risk of post-stroke infection: a pilot study. *Clin. Proteomics* **14**, 27 (2017).
44. Hart, R. G. et al. Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol.* **13**, 429–438 (2014).
45. Llombart, V. et al. B-type natriuretic peptides help in cardioembolic stroke diagnosis: pooled data meta-analysis. *Stroke* **46**, 1187–1195 (2015).
46. Bai, J., Sun, H., Xie, L., Zhu, Y. & Feng, Y. Detection of cardioembolic stroke with B-type natriuretic peptide or N-terminal pro-BNP: a comparative diagnostic meta-analysis. *Int. J. Neurosci.* **128**, 1100–1108 (2018).
47. Santamarina, E. et al. Biomarker level improves the diagnosis of embolic source in ischemic stroke of unknown origin. *J. Neurol.* **259**, 2538–2545 (2012).
48. Kamel, H. et al. The AtRial Cardiopathy and Antithrombotic Drugs In prevention After cryptogenic stroke randomized trial: rationale and methods. *Int. J. Stroke* **14**, 207–214 (2019).
49. Engdahl, J. et al. Stepwise mass screening for atrial fibrillation using N-terminal pro B-type natriuretic peptide: the STROKESTOP II study design. *Europace* **19**, 297–302 (2017).
50. Ward, G., Simpson, A., Boscato, L. & Hickman, P. E. The investigation of interferences in immunoassay. *Clin. Biochem.* **50**, 1306–1311 (2017).
51. Andreasson, U. et al. A practical guide to immunoassay method validation. *Front. Neurol.* **6**, 179 (2015).
52. Schimmel, H., Zegers, I. & Emons, H. Standardization of protein biomarker measurements: is it feasible? *Scand. J. Clin. Lab. Invest. Suppl.* **70**, 27–33 (2010).
53. Harpaz, D., Eltzov, E., Seet, R. C. S., Marks, R. S. & Tok, A. I. Y. Point-of-care-testing in acute stroke management: an unmet need ripe for technological harvest. *Biosensors* **7**, 30 (2017).
54. Cuadrado, E. et al. Vascular MMP-9/TIMP-2 and neuronal MMP-10 up-regulation in human brain after stroke: a combined laser microdissection and protein array study. *J. Proteome Res.* **8**, 3191–3197 (2009).
55. Cuadrado, E. et al. The proteome of human brain after ischemic stroke. *J. Neuropathol. Exp. Neurol.* **69**, 1105–1115 (2010).
This study is the first to describe the whole brain proteome after ischaemic stroke.
56. García-Berrocoso, T. et al. From brain to blood: new biomarkers for ischemic stroke prognosis. *J. Proteom.* **94**, 138–148 (2013).
57. Yuan, D., Liu, C. & Hu, B. Dysfunction of membrane trafficking leads to ischemia–reperfusion injury after transient cerebral ischemia. *Transl Stroke Res.* **9**, 215–222 (2018).
58. García-Berrocoso, T. et al. Single cell immuno-laser microdissection coupled to label-free proteomics to reveal the proteotypes of human brain cells after ischemia. *Mol. Cell. Proteom.* **17**, 175–189 (2018).
59. Dayon, L. et al. Brain extracellular fluid protein changes in acute stroke patients. *J. Proteome Res.* **10**, 1043–1051 (2011).
60. Engelborghs, S. et al. Consensus guidelines for lumbar puncture in patients with neurological diseases. *Alzheimers Dement.* **8**, 111–126 (2017).
61. Domingues, R., Bruniera, G., Brunale, F., Mangueira, C. & Senne, C. Lumbar puncture in patients using anticoagulants and antiplatelet agents. *Arq. Neuropsiquiatr.* **74**, 679–86 (2016).
62. Zimmermann-Ivol, C. G. et al. Fatty acid binding protein as a serum marker for the early diagnosis of stroke. *Mol. Cell. Proteomics* **3**, 66–72 (2004).
63. Lescuyer, P. et al. Identification of post-mortem cerebrospinal fluid proteins as potential biomarkers of ischemia and neurodegeneration. *Proteomics* **4**, 2234–2241 (2004).
64. Dayon, L. et al. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal. Chem.* **8**, 2921–2931 (2008).
65. Simats, A. et al. Characterization of the rat cerebrospinal fluid proteome following acute cerebral ischemia using an aptamer-based proteomic technology. *Sci. Rep.* **8**, 7899 (2018).
66. Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867 (2002).
67. Agouni, A. et al. There is selective increase in pro-thrombotic circulating extracellular vesicles in acute ischemic stroke and transient ischemic attack: a study of patients from the Middle East and Southeast Asia. *Front. Neurol.* **10**, 251 (2019).
68. Ciccocioppo, F., Lanuti, P., Centonze, D., Miscia, S. & Marchisio, M. The link among neurological diseases: extracellular vesicles as a possible brain injury footprint. *Neurosignals* **27**, 25–39 (2019).
69. Turck, N. et al. A multiparameter panel method for outcome prediction following aneurysmal subarachnoid hemorrhage. *Intensive Care Med.* **36**, 107–115 (2010).
70. Woo, D. et al. Top research priorities for stroke genetics. *Lancet Neurol.* **17**, 663–665 (2018).
71. Psaty, B. M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (charge) consortium design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc.Genet.* **2**, 73–80 (2009).
72. Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
This study is the latest and largest GWAS of stroke, including 72,147 patients with stroke, that identified 32 loci associated with stroke.
73. Malik, R. et al. Genome-wide meta-analysis identifies 3 novel loci associated with stroke. *Ann. Neurol.* **84**, 934–939 (2018).
74. Okazaki, S. et al. Moyamoya disease susceptibility variant RNF213 p.R4810K increases the risk of ischemic stroke attributable to large-artery atherosclerosis. *Circulation* **139**, 295–298 (2019).
75. Kubo, M. et al. A nonsynonymous SNP in PRKCH (protein kinase C η) increases the risk of cerebral infarction. *Nat. Genet.* **39**, 212–217 (2007).
76. Woo, D. et al. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am. J. Hum. Genet.* **94**, 511–521 (2014).
77. Traylor, M. et al. A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. *PLoS Genet.* **10**, e1004469 (2014).
78. Traylor, M. et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol.* **11**, 951–962 (2012).
79. Debette, S. et al. Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. *Nat. Genet.* **47**, 78–83 (2014).
80. Malik, R. et al. ISGC Analysis Group, METASTROKE collaboration, Wellcome Trust Case Control Consortium 2 (WTCCC2), NINDS Stroke Genetics Network (SiGN). Low-frequency and common genetic variation in ischemic stroke: the METASTROKE collaboration. *Neurology* **86**, 1217–1226 (2016).
81. Bellenguez, C. et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat. Genet.* **44**, 328–333 (2012).
82. Holliday, E. G. et al. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat. Genet.* **44**, 1147–1151 (2012).
83. 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**, 678–85 (2014).
84. Ikram, M. A. et al. Genomewide association studies of stroke. *N. Engl. J. Med.* **360**, 1718–1728 (2009).
85. Biffi, A. et al. APOE genotype and extent of bleeding and outcome in lobar intracerebral haemorrhage: a genetic association study. *Lancet Neurol.* **10**, 702–709 (2011).
86. Cheng, Y. C. et al. Genome-wide association analysis of young-onset stroke identifies a locus on chromosome 10q25 near HABP2. *Stroke* **47**, 307–316 (2016).
87. Gretarsdottir, S. et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann. Neurol.* **64**, 402–409 (2008).
88. Carty, C. L. et al. Meta-analysis of genome-wide association studies identifies genetic risk factors for stroke in African Americans. *Stroke* **46**, 2063–2068 (2015).
89. Gudbjartsson, D. F. et al. A sequence variant in ZFHX3 on 16q22 associates with a trial fibrillation and ischemic stroke. *Nat. Genet.* **41**, 876–878 (2009).
90. Rosand, J. et al. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol.* **15**, 174–184 (2016).
91. Chauhan, G. et al. Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **15**, 695–707 (2016).
92. Verhaaren, B. F. J. et al. Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circ. Cardiovasc. Genet.* **8**, 398–409 (2015).
93. Cole, J. W. et al. Genetics of the thrombomodulin–endothelial cell protein C receptor system and the risk of early-onset ischemic stroke. *PLoS One* **13**, e0206554 (2018).
94. Rannikmäe, K. et al. Common variation in COL4A1/COL4A2 is associated with sporadic cerebral small vessel disease. *Neurology* **84**, 918–926 (2015).
95. Gupta, R. M. et al. A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell* **170**, 522–533 (2017).
96. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
97. Söderholm, M. et al. International Stroke Genetics Consortium, the NINDS-SiGN Consortium, and the Genetics of Ischaemic Stroke Functional Outcome (GISCOME) Network. Genome-wide association meta-analysis of functional outcome after ischemic stroke. *Neurology* **92**, e1271–e1283 (2019).
This meta-analysis describes genetic variants associated with functional outcomes 3 months after stroke.
98. Marini, S. et al. 17p12 influences hematoma volume and outcome in spontaneous intracerebral hemorrhage. *Stroke* **49**, 1618–1625 (2018).
99. Dorothea Pfeiffer, Bowang et al. Genetic imbalance is associated with functional outcome after ischemic stroke. *Stroke* **50**, 298–304 (2019).
100. Ibanez, L. et al. Overlap in the genetic architecture of stroke risk, early neurological changes, and cardiovascular risk factors. *Stroke* **50**, 1339–1345 (2019).
101. Schlunk, F. & Greenberg, S. M. The pathophysiology of intracerebral hemorrhage formation and expansion. *Transl Stroke Res.* **6**, 257–263 (2015).
102. Pulit, S. L. et al. Atrial fibrillation genetic risk differentiates cardioembolic stroke from other stroke subtypes. *Neurol. Genet.* **4**, e293 (2018).

103. Dick, K. J. et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet* **383**, 1990–1998 (2014).
104. Liang, L. et al. An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature* **520**, 670–674 (2015).
105. Gallego-Fabrega, C. et al. PPM1A methylation is associated with vascular recurrence in aspirin-treated patients. *Stroke* **47**, 1926–1929 (2016).
106. Soriano-Tárraga, C. et al. Global DNA methylation of ischemic stroke subtypes. *PLoS One* **9**, e96543 (2014).
107. Soriano-Tárraga, C. et al. Biological age is better than chronological as predictor of 3-month outcome in ischemic stroke. *Neurology* **89**, 830–836 (2017).
108. Mishra, A. et al. Association of variants in HTRA1 and NOTCH3 with MRI-defined extremes of cerebral small vessel disease in older subjects. *Brain* **142**, 1009–1023 (2019).
109. Raffield, L. M. et al. Coagulation factor VIII: relationship to cardiovascular disease risk and whole genome sequence and epigenome-wide analysis in African Americans. *J. Thromb. Haemost.* https://doi.org/10.1111/jth.14741 (2020).
110. Wojcik, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
111. Musunuru, K. et al. The expressed genome in cardiovascular diseases and stroke: refinement, diagnosis, and prediction: a scientific statement from the American Heart Association. *Circ. Cardiovasc. Genet.* **10**, e000037 (2017).
112. Moore, D. F. et al. Using peripheral blood mononuclear cells to determine a gene expression profile of acute ischemic stroke: a pilot investigation. *Circulation* **11**, 212–221 (2005).
113. Stamova, B. et al. Gene expression profiling of blood for the prediction of ischemic stroke. *Stroke* **41**, 2171–2177 (2010).
114. Stamova, B. et al. The intracerebral hemorrhage blood transcriptome in humans differs from the ischemic stroke and vascular risk factor control blood transcriptomes. *J. Cereb. Blood Flow Metab.* **39**, 1818–1835 (2018).
115. Jickling, G. C. et al. Signatures of cardioembolic and large-vessel ischemic stroke. *Ann. Neurol.* **68**, 681–692 (2010).
116. Xu, H. et al. Gene expression in peripheral blood differs after cardioembolic compared with large-vessel atherosclerotic stroke: biomarkers for the etiology of ischemic stroke. *J. Cereb. Blood Flow. Metab.* **28**, 1320–1328 (2008).
117. Jickling, G. C. et al. Profiles of lacunar and nonlacunar stroke. *Ann. Neurol.* **70**, 477–485 (2011).
118. Jickling, G. C. et al. Prediction of cardioembolic, arterial, and lacunar causes of cryptogenic stroke by gene expression and infarct location. *Stroke* **43**, 2036–2041 (2012).
119. Jickling, G. C. et al. Leukocyte response is regulated by microRNA let7i in patients with acute ischemic stroke. *Neurology* **87**, 2198–2205 (2016).
120. Gilles, M. E. & Slack, F. J. Let-7 microRNA as a potential therapeutic target with implications for immunotherapy. *Expert Opin. Ther. Targets* **22**, 929–939 (2018).
121. Tiedt, S. et al. RNA-seq identifies circulating MIR-125a-5p, MIR-125b-5p, and MIR-143-3p as potential biomarkers for acute ischemic stroke. *Circ. Res.* **121**, 970–980 (2017).
122. Eyileten, C. et al. MicroRNAs as diagnostic and prognostic biomarkers in ischemic stroke — a comprehensive review and bioinformatic analysis. *Cells* **7**, 249 (2018).
123. Tiedt, S. & D., M. Role of non-coding RNAs in stroke. *Stroke* **49**, 3098–3106 (2018).
124. Jickling, G. C. et al. MicroRNA expression in peripheral blood cells following acute ischemic stroke and their predicted gene targets. *PLoS One* **9**, e99283 (2014).
125. Huang, S. et al. Identification of blood let-7e-5p as a biomarker for ischemic stroke. *PLoS One* **11**, e0163951 (2016).
126. Tan, K. S. et al. Expression profile of microRNAs in young stroke patients. *PLoS One* **4**, e7689 (2009).
127. Sepramaniam, S. et al. Circulating microRNAs as biomarkers of acute stroke. *Int. J. Mol. Sci.* **15**, 1418–1432 (2014).
128. Dykstra-Aiello, C. et al. Altered expression of long noncoding RNAs in blood after ischemic stroke and proximity to putative stroke risk loci. *Stroke* **47**, 2896–2903 (2016).
129. Deng, Q.-W. et al. Differential long noncoding RNA expressions in peripheral blood mononuclear cells for detection of acute ischemic stroke. *Clin. Sci.* **132**, 159701501614 (2018).
130. Wang, W. et al. Integrated analysis of lncRNA-mRNA co-expression profiles in patients with Moyamoya disease. *Sci. Rep.* **7**, 42421 (2017).
131. Holdt, L. M. & Teupser, D. Long Noncoding R. N. A. ANRIL: Lnc-ing genetic variation at the chromosome 9p21 locus to molecular mechanisms of atherosclerosis. *Front. Cardiovasc. Med.* **5**, 145 (2018).
132. Wang, J. et al. Predictive value of long noncoding RNA ZFAS1 in patients with ischemic stroke. *Clin. Exp. Hypertens.* **41**, 615–621 (2018).
133. Wishart, D. S. et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
134. Shah, S. H., Kraus, W. E. & Newgard, C. B. Metabolomic profiling for the identification of novel biomarkers and mechanisms related to common cardiovascular diseases form and function. *Circulation* **126**, 1110–1120 (2012).
135. McGarrah, R. W., Crown, S. B., Zhang, G. F., Shah, S. H. & Newgard, C. B. Cardiovascular metabolomics. *Circ. Res.* **122**, 1238–1258 (2018).
136. Marklund, M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Fatty Acids and Outcomes Research Consortium (FORCE). Biomarkers of dietary omega-6 fatty acids and incident cardiovascular disease and mortality. *Circulation* **139**, 2422–2436 (2019).
137. Stegemann, C. et al. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. *Circulation* **129**, 1821–1831 (2014).
138. Sun, D. et al. A prospective study of serum metabolites and risk of ischemic stroke. *Neurology* **92**, e1890–e1898 (2019).
139. Hu, Z. et al. Rapid and sensitive differentiating ischemic and hemorrhagic strokes by dried blood spot based direct injection mass spectrometry metabolomics analysis. *J. Clin. Lab. Anal.* **30**, 823–830 (2016).
140. Sheth, S. A., Iavarone, A. T., Liebeskind, D. S., Won, S. J. & Swanson, R. A. Targeted lipid profiling discovers plasma biomarkers of acute brain injury. *PLoS One* **10**, e0129735 (2015).
141. Kimberly, W. T., Wang, Y., Pham, L., Furie, K. L. & Gerszten, R. E. Metabolite profiling identifies a branched chain amino acid signature in acute cardioembolic stroke. *Stroke* **44**, 1389–1395 (2013).
142. Choi, J. Y. et al. High free fatty acid level is associated with recurrent stroke in cardioembolic stroke patients. *Neurology* **82**, 1142–1148 (2014).
143. Adams, H. P. 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* **24**, 35–41 (1993).
144. Ay, H. et al. An evidence-based causative classification system for acute ischemic stroke. *Ann. Neurol.* **58**, 688–697 (2005).
145. Nelson, S. E., Ament, Z., Wolcott, Z., Gerszten, R. E. & Kimberly, W. T. Succinate links atrial dysfunction and cardioembolic stroke. *Neurology* **92**, e802–e810 (2019).
This study links energy metabolites, such as succinate, to cardioembolic stroke and structural cardiac abnormalities for the first time, thereby supporting the abnormal atrial substrate model of cardioembolic stroke.
146. Kamel, H., Okin, P. M., Elkind, M. S. V. & Iadecola, C. Atrial fibrillation and mechanisms of stroke: time for a new model. *Stroke* **47**, 895–900 (2016).
147. Kennedy, J. et al. Rapid screening and identification of novel psychoactive substances using PaperSpray interfaced to high resolution mass spectrometry. *Clin. Mass Spectrom.* **1**, 3–10 (2016).
148. Land, W. H. et al. Partial least squares (PLS) applied to medical bioinformatics. *Procedia Comput. Sci.* **6**, 273–278 (2011).
149. de Tayrac, M., Lê, S., Aubry, M., Mosser, J. & Husson, F. Simultaneous analysis of distinct omics data sets with integration of biological knowledge: multiple factor analysis approach. *BMC Genomics* **10**, 32 (2009).
150. Cisek, K., Krochmal, M., Klein, J. & Mischak, H. The application of multi-omics and systems biology to identify therapeutic targets in chronic kidney disease. *Nephrol. Dial. Transpl.* **31**, 2003–2011 (2016).
151. Hopewell, J. C. & Clarke, R. Emerging risk factors for stroke: what have we learned from Mendelian randomization studies? *Stroke* **47**, 1673–1678 (2016).
152. Larsson, S. C. et al. Serum 25-hydroxyvitamin D concentrations and ischemic stroke and its subtypes. A Mendelian randomization study. *Stroke* **49**, 2508–2511 (2018).
153. Larsson, S. C., Traylor, M. & Markus, H. S. Homocysteine and small vessel stroke: a Mendelian randomization analysis. *Ann. Neurol.* **85**, 495–501 (2019).
154. van der Laan, S. W. et al. Cystatin C and cardiovascular disease: a Mendelian randomization study. *J. Am. Coll. Cardiol.* **68**, 934–945 (2016).
155. Georgakis, M. K. et al. Genetically determined levels of circulating cytokines and risk of stroke. *Circulation* **139**, 256–268 (2019).
156. Marios, K. et al. Circulating monocyte chemoattractant protein-1 and risk of stroke: meta-analysis of population-based studies involving 17 180 individuals. *Circ. Res.* **125**, 773–782 (2019).
157. Sun, L. et al. Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults. *Nat. Med.* **25**, 569–574 (2019).
158. Hartwig, F. P., Davies, N. M., Hemani, G. & Davey Smith, G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int. J. Epidemiol.* **45**, 1717–1726 (2016).
159. Ference, B. A. et al. Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. *N. Engl. J. Med.* **375**, 2144–2153 (2016).
160. Mokry, L. E., Ahmad, O., Forgetta, V., Thanassoulis, G. & Richards, J. B. Mendelian randomisation applied to drug development in cardiovascular disease: a review. *J. Med. Genet.* **52**, 71–79 (2015).
161. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
162. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
163. Bayat Mokhtari, R. et al. Combination therapy in combating cancer. *Oncotarget* **8**, 38022–38043 (2017).
164. Palmer, A. C. & Sorger, P. K. Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell* **14**, 1678–1691 (2017).
165. West, L. et al. A novel classification of lung cancer into molecular subtypes. *PLoS One* **7**, e31906 (2012).
166. Collisson, E. A. et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* **17**, 500–503 (2011).
167. Higdon, R. et al. The promise of multi-omics and clinical data integration to identify and target personalized healthcare approaches in autism spectrum disorders. *OMICS* **19**, 197–208 (2015).
168. Sertbas, M. & Ulgen, K. O. Unlocking human brain metabolism by genome-scale and multiomics metabolic models: relevance for neurology research, health, and disease. *OMICS* **22**, 455–467 (2018).
169. Traylor, M. et al. Genetics of stroke in a UK African ancestry case–control study: South London Ethnicity and Stroke Study. *Neurol. Genet.* **3**, e142 (2017).
170. Zakai, N. A. et al. D-dimer and the risk of stroke and coronary heart disease. The REasons for Geographic and Racial Differences in Stroke (REGARDS) Study. *Thromb Haemost* **117**, 618–624 (2017).
171. Zhan, X. et al. Transient ischemic attacks characterized by RNA profiles in blood. *Neurology* **77**, 1718–1724 (2011).
172. Jickling, G. C. et al. Ischemic transient neurological events identified by immune response to cerebral ischemia. *Stroke* **43**, 1006–1012 (2012).
This study describes the role of RNA in predicting the cause of stroke when the cause remains unclear or cryptogenic.
173. Azghandi, S. et al. Deficiency of the stroke relevant HDAC9 gene attenuates atherosclerosis in accord with allele-specific effects at 7p21.1. *Stroke* **46**, 197–202 (2015).

174. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
175. Sakaue, S. & Okada, Y. GREP: genome for REPositioning drugs. *Bioinformatics* **35**, 3821–3823 (2019).

Acknowledgements
The Neurovascular Research Laboratory acknowledges funding from grants PI15/00354 and PI18/00804 from Fondo de Investigaciones Sanitarias of the Instituto de Salud Carlos III (co-financed by the European Regional Development Fund, FEDER). The Neurovascular Research Laboratory is also a member of the Spanish stroke research network INVICTUS+ (RD16/0019). L.R. is supported by a predoctoral fellowship grant from the Instituto de Salud Carlos III (IFI17/00012). A.B. is supported by a Juan Rodes research grant from Instituto de Salud Carlos III (JR16/00008).

Author contributions
All authors contributed to the development of the manuscript, wrote sections of the manuscript, approved the final version and are responsible for the content.

Competing interests
Members of the neurovascular research laboratory (A.S., A.B., J.M. and L.R.) are inventors of a family of patents for biomarkers to differentiate ischaemic from haemorrhagic stroke, to predict stroke outcome and to establish stroke aetiology. J.-C.S. and J.M. are co-founders of ABCDx, a spin-off company of the University of Geneva (http://www.abcdx.ch).

Peer review information
Nature Reviews Neurology thanks J. Meschia, B. Worrall and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

RELATED LINKS
ISGC: http://www.strokegenetics.org