Analyst

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Analyst*, 2020, **145**, 1695

Global metabolomics analysis of serum from humans at risk of thrombotic stroke†

Adnan Khan, ‡[a] Mal-Soon Shin, ‡[b] Sun Ha Jee [c] and Youngja H. Park *[a]

We aimed to determine the serum concentrations of altered compounds to understand the changes in metabolism and pathophysiology that occur prior to thrombotic stroke. In this prospective cohort study, high-resolution metabolomics (HRM) was employed to analyze serum samples obtained from patients at risk of stroke (n = 99) and non-risk controls (n = 301). Partial least-squares discriminant analysis (PLS-DA), along with univariate analysis using a false discovery rate (FDR) of q = 0.05 were employed to identify the discriminant metabolic profiles and to determine significantly different metabolites between healthy control and stroke risk groups. PLS-DA satisfactorily separated the stroke risk sera from control sera. Additionally, these discriminant metabolic profiles were not related to hypertension, smoking, diabetes mellitus, or insulin sensitivity. A group of 35 metabolites, most of them amino acids, that were capable of discriminating stroke risk sera from controls were identified using untargeted metabolomics. Further, the targeted metabolomics approach confirmed that the quantified concentrations of L-tryptophan, 3-methoxytyramine, methionine, homocysteinesulfinic acid, cysteine, isoleucine, carnitine, arginine, linoleic acid, and sphingosine were specifically elevated in the sera of patients who were later diagnosed with stroke. Our untargeted and targeted metabolomics approaches support investigating these compounds as novel biomarkers for early and non-invasive detection of thrombotic stroke.

Received 11th October 2019,
Accepted 16th December 2019

DOI: 10.1039/c9an02032b

rsc.li/analyst

1. Introduction

Stroke is the second leading cause of death worldwide.[1] Stroke is either caused by cerebral ischemia due to thrombotic occlusion or hemorrhage causing transient or permanent neurological disorders and constitutes a heavy burden on society.[2] Up to 80% of strokes are ischemic or thrombotic strokes, while the remaining 20% are primarily due to hemorrhage.[1,3] A reduction in cerebral blood flow due to thrombosis causes an energy crisis and lack of oxygen.[3,4] Such energy depletion in turn leads to a cascade of molecular events, such as the release of excitatory glutamate, oxidative stress, and an inflammatory response.[3] During these molecular events, numerous biomarkers are released into the peripheral blood, which provide an opportunity to utilize metabolomics research to elucidate new clinical prognostic biomarkers for ischemic lesions.[3]

In the past 10 years, a great deal of research has been performed in the stroke research field; however, the prevalence of stroke has not been markedly reduced.[5] A limited number of therapeutic options are commercially available for stroke, and effective therapies are desperately needed. The high mortality of stroke patients is in part due to the silent asymptomatic progression of the disease.[6] By the time stroke becomes symptomatic, it is already at an irreversible stage, and it has been estimated that at this point, there are already 10 silent infarcts occupying the brain.[7] Therefore, this study used high resolution metabolomics (HRM) to evaluate people at risk for stroke in order to investigate early markers of endothelial and vascular dysfunction in an attempt to identify a disease that has yet to become symptomatic.

The metabolite biomarkers present a footprint of biological processes that may expose implications in pathophysiologic pathways during pathological conditions.[8] A minor change in the expression level of a gene or a protein due to a physiological or pathological condition can cause a significant change in the levels of metabolites.[8] HRM can allow for the extraction of global metabolic pathway information from tens of thousands of metabolites present in biological samples.[8] Therefore, metabolomics offers a promising approach to identifying biomarkers or a panel of biomarkers that could serve as indicators of a specific condition. In our previous study, we applied HRM, coupled with liquid chromatography-mass spectrometry

[a]*Metabolomics Laboratory, Korea University College of Pharmacy, Sejong 30019, Republic of Korea. E-mail: yjhwang@korea.ac.kr; Fax: +82-44-860-1606; Tel: +82-44-860-1621*
[b]*School of Global Sport Studies, Korea University, Sejong 30019, Republic of Korea*
[c]*Department of Epidemiology and Health Promotion and Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul 03722, Republic of Korea*

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c9an02032b

‡These authors contributed equally.

(LC-MS) in a prospective cohort study, using serum samples obtained from patients at risk of thrombotic stroke and non-risk individuals.[3] Our study identified lysine and valine catabolites, along with ubiquinone and homocysteine sulfinic acid as potential metabolites that could discriminate the thrombotic stroke risk group from the control group.[3] However, our previous study did not employ targeted metabolomics profiling using liquid chromatography with tandem mass spectrometry (LC-MS/MS).[3] In addition, in response to any successful metabolomics study, it is important to further employ large validation studies with suitable control cohorts in order to remove any potential bias.[9]

Therefore, this study applied both the untargeted and targeted HRM techniques in a large prospective cohort study using sera samples in an attempt to validate our previously published results and to further investigate the global metabolic perturbation in asymptotic stroke. Finally, this study aimed to identify a panel of predictive biomarkers of thrombotic stroke to ease the burden of stroke diagnosis and prognosis, which may further reduce the prevalence of stroke in the future and provide a new mechanistic insight for improving the understanding of the pathophysiologic basis for thrombotic stroke in the developmental phase.

2. Method and materials

2.1 Sample collection

The study protocol was approved by the Korea University Institutional Review Board (KU-IRB-15-19-A-1) and Institutional Review Board of Human Research of Yonsei University. Serum samples used in this study were obtained from the Korean Cancer Prevention Study-II (KCPS-II) Biobank, Seoul and Gyeonggi, South Korea.[3] The study cohort was comprised of 156 701 participants (94 840 men and 61 861 women), and data collection began in 2004. The participants underwent routine health assessments, provided blood samples, and gave informed consent for long-term prospective follow-up.[10] Sampling was conducted according to the protocols approved by South Korea's Bioethics and Safety Act No. 9100. Blood was obtained from 400 people who consented to testing for research purposes at 11 comprehensive examination centers located in Seoul and Gyeonggi Province in South Korea, between 2004 and 2015. Blood samples were taken after overnight fasting. Median blood collection was achieved by the end of 2008, with 90% of blood collections between mid-2005 and the end of 2008. The collected serum samples were frozen at −80 °C until analysis and freeze–thaw cycles were avoided to minimize release of intracellular metabolites. The incidence of stroke cases were determined by the hospital admission discharge records from 2005 to 2015 (median follow-up duration, 8.0 years). These outcome data were obtained from health insurance claims provided by the National Health Insurance Service. Glucose, total cholesterol, triglycerides, high-density lipoprotein cholesterol (HDL-c), low-density lipoprotein cholesterol (LDL-c), and other biomarkers were measured in the hospital laboratory by a COBAS INTEGRA 800 and a 7600 Analyzer (Hitachi, Tokyo, Japan). The participants were then classified into two groups based on their diagnosis as stroke or non-stroke patients. Among them, 301 subjects who were not diagnosed with thrombotic stroke later were categorized as the control group, while 99 subjects, who were later diagnosed with thrombotic stroke, were categorized as the stroke risk group. Details such as sex, age, fasting blood sugar (FBS), total cholesterol, triglycerides, high density lipoproteins, low density lipoproteins, blood pressure, smoking, diabetes mellitus, and insulin sensitivity status of the control group and stroke risk patients are provided in Table 1.

Table 1 Demographic characteristics of subjects

	Control	Stroke
n	301	99
Male (*n*)	177 (58.8%)	72 (72.7%)
Female (*n*)	124 (41.2%)	27 (27.3%)
Age (year)	42.1 ± 7.7	52.8 ± 11.8
Fasting blood sugar (mg dL^{-1})	90.2 ± 15.7	98.5 ± 25.2*
Total cholesterol (mg dL^{-1})	189.8 ± 32.2	193.7 ± 31.0
Triglycerides (mg dL^{-1})	126.9 ± 74.6	161.9 ± 98.5*
High density lipoproteins (mg dL^{-1})	52.4 ± 11.3	50.5 ± 10.7
Low density lipoproteins (mg dL^{-1})	115.6 ± 31.4	115.2 ± 31.9
Systolic blood pressure (mmHg)	118.8 ± 14.6	126.8 ± 17.4*
Diastolic blood pressure (mmHg)	78.2 ± 1.1	80.1 ± 10.7*
Smokers	82 (27.2%)	36 (36.4%)
Diabetes mellitus	20 (6.6%)	19 (19.2%)
Insulin-sensitive	138 (45.8%)	64 (64.6%)

All clinical values are expressed as mean ± SD. * represents $p \leq 0.05$.

2.2 Chemicals and reagents used for metabolomics analysis

High-performance liquid chromatography (HPLC)-grade water was purchased from J.T. Baker (Phillipsburg, NJ, USA) and acetonitrile was from Tedia (Fair Lawn, NJ, USA). Formic acid was purchased from Fluka (St Louis, MO, USA). All chemicals and reagents were stored at appropriate temperatures and conditions. Standard solutions and serum samples were stored at −80 °C.[11]

2.3 Sample preparation for metabolite extraction

Sample preparation was performed as discussed previously.[12] Briefly 50 μL aliquots of sera obtained from stroke risk patients and healthy controls were first treated with 195 μL of acetonitrile and 5 μL of a mixture of three stable isotope standards ([3-methyl-13C]-caffeine, [dimethyl-D6]-*N*,*N*-diethyl-*M*-toluamide, and [13C5, 15N]-L-methionine) (1 : 4, v/v).[12] The samples were than vortexed and centrifuged at 13 000 rpm at 4 °C for 10 min for protein precipitation and metabolite extraction. The supernatants containing the polar metabolites were collected for LC-MS/MS analysis. Samples were stored at −80 °C until use.

2.4 Analysis of metabolites by LC-MS/MS

An Ultra Performance Liquid Chromatography system (Agilent 1260 Infinity Quaternary) coupled with an Agilent LC-MS/MS

Q-TOF 6550 mass spectrometer was used for the metabolomics profiling. The samples were analyzed using C18 Synchronis aQ (1.9 μm, 100 × 2.1 mm; Thermo Fisher Scientific, Inc., Waltham, MA, USA). The column and autosampler temperature were maintained at 45 °C and 10 °C, respectively. Solvent A consisted of 0.1% formic acid in water, and solvent B consisted of 0.1% formic acid in acetonitrile. The injection volume and flow rate were 5 μL and 0.4 mL min^{-1}, respectively. The HPLC gradient was programmed as follows: 95% water for 1 min, a linear decrease to 55% water over 8 min, a descending gradient to 10% water over 3 min, a 1.5 min hold and return to 95% water over 0.1 min. The electrospray ionization detector was operated with a curtain gas of 35 psi at 250 °C, supplied at 14 mL min^{-1}, and sheath gas temperature of 250 °C, supplied at a flow rate of 11 mL min^{-1}. Detection of the mass/charge ratio (m/z) of ions was set from 50 to 1000, with a resolution of 20 000 over 15 min. All samples were run in triplicate, and data for each ionization technique were acquired in positive ion mode.[13]

2.5 Metabolic profiling

To identify the metabolic features that discriminate the stroke risk patients from controls, multivariate and univariate analyses were performed. The apLCMS was first used to analyze all the features of the samples for subsequent statistical analyses.[14] The apLCMS provided 7197 m/z (mass/charge ratio) values within a range of ions set from 50 to 1000 from mass spectral data.[14] False discovery rates (FDRs) were calculated to reduce the incidence of false-positives, and Manhattan plots were constructed using MetaboAnalyst 4.0 (http://www.metaboanalyst.ca) to identify metabolites, whose levels were significantly different between stroke risk sera and control sera, according to Student's t-test. Supervised multivariate partial least squares-discriminant analysis (PLS-DA) was performed using SIMCA 14.1 (Umetrics AB, Umeå, Sweden). To ensure the quality of the PLS-DA models and to avoid the risk of overfitting, 7-fold cross-validation (CV) was applied with six principal components, as 7-fold cross-validation is the default SIMCA cross-validation procedure.[15] Two parameters: R^2 (goodness of fit) and Q^2 (goodness of prediction) were evaluated for each PLS-DA model.[15] The performance of PLS-DA models was also validated by a permutation test (20 times) using six components.[15] PLS-DA analyses were performed as per the schemes of analysis shown in Fig. 1; sera from control patients were compared with sera from stroke risk patients, collectively. Further, to observe the effect of smoking, hypertension (HTN), diabetes mellitus (DM), and insulin sensitivity (IS) on metabolic disturbances, the smokers, hypertensive, DM, and IS control sera were compared with the smokers, hypertensive, DM, and IS stroke risk sera using PLS-DA, respectively.

2.6 Metabolic pathway analyses

The metabolites with an FDR-adjusted p value <0.05 were considered as significantly different metabolite features between the stroke risk group and control group in Manhattan plots.

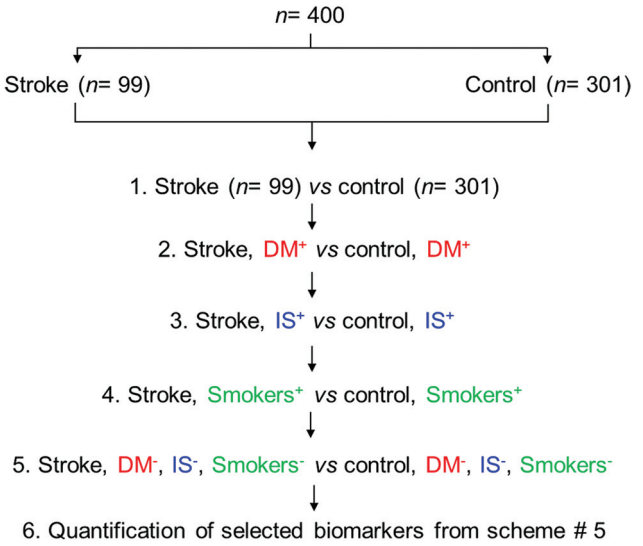

Fig. 1 Flow chart of the subjects used in this prospective cohort study and metabolomics. HTN represents hypertension, Sm represents smokers, DM represents diabetes mellitus, is represents insulin sensitivity.

These metabolites were considered important in the identification of potential biomarkers and were subsequently fed into xMSannotator[16] and the Metlin Mass Spectrometry Database (METLIN) (https://metlin.scripps.edu). The following positive ion adducts were used during annotation: $[M + H]^+$, $[M + NH_4]^+$, $[M + Na]^+$, $[M + H - H_2O]^+$, and $[M + K]^+$, with a confidence limit of 10 ppm to increase sensitivity in the identification of compounds and to filter out irrelevant compounds.[12] xMSannotator and METLIN (https://metlin.scripps.edu) provided the KEGG numbers for each m/z.[16] For identification of potentially altered metabolic pathways in control *versus* stroke risk patients, the recorded KEGG numbers served as input for the human metabolomics pathway in the Kyoto Encyclopedia of Genes and Genome database (http://www.kegg.jp). Compounds with significant changes (p-value <0.05) were subsequently considered important for identification as potential biomarkers related to metabolic effects caused by infract development.

2.7 Targeted metabolite profiling

For the identification and quantification of metabolites, the reference standards were purchased from Sigma Chemical Co. (St Louis, MO, USA). The standards were weighed accurately, dissolved in methanol/water, as per the instructions for the materials, and stored at 4 °C. The non-smokers, non-hypertensive, non-DM, and non-IS serum samples of control and stroke risk patients were chosen for the quantification of the potential metabolic biomarkers. Serum samples were treated with acetonitrile (1 : 4, v/v) and centrifuged to precipitate proteins. Tandem mass spectrometry (MS/MS) data were acquired in the positive ion mode using an Agilent 6490A Triple Quad Mass Spectrometer (Agilent Technologies, Inc.) with an accompanying ESI interface.[12] The standards and serum samples were

first scanned in the mass range (m/z) 50–1000. Collision energy of 0, 5, 10, 15, and 20 V was then used to produce the highly abundant fragment ions of the putative metabolites during product-ion analysis in the positive mode (Table S1†). Chromatography was performed on a C18 (100 × 2.1 mm) column (Higgins Analytical, Inc., Mountain View, CA, USA), at a flow rate of 0.4 mL min^{-1}. Concentrations of identified metabolites in serum samples from control or stroke risk patients were quantified by making the calibration curve of each standard compound with at least eight appropriate concentration levels.[12] The limit of detection (LOD) and limit of quantification (LOQ) under the present chromatographic conditions were determined at a signal-to-noise (S/N) ratio of 3 and 10, respectively.[12] The analyses were performed in triplicate, and data are presented as the mean ± SEM. The concentrations of targeted metabolites were calculated by reference to the peak areas of the external standards within the range of LOD and LOQ.

2.8 Statistical analysis using GraphPad

Putative identities were analyzed using the GraphPad Prism software (v. 5.03; La Jolla, CA, USA) for measurement of their relative intensities in each group. Data are presented as the means ± SEM and differences, with p-values <0.05 considered statistically significant.

3. Results

3.1 Subject characteristics

The subjects in this study were from a cohort of stroke risk patients and a control group living in South Korea. Metabolomics analysis was performed on a total of 99 stroke risk patients and 301 control subjects. No statistical differences were observed in the age, total cholesterol (mg dL^{-1}), and high density lipoproteins (mg dL^{-1}), based on Student's t-test among the two groups (Table 1). However, levels of fasting blood sugar (mg dL^{-1}), triglycerides (mg dL^{-1}), and systolic blood pressure (mmHg) as well as diastolic blood pressure (mmHg) were significantly higher in the stroke risk patients compared to the control subjects, as shown in Table 1.

3.2 Differential metabolome of stroke risk sera in comparison with control sera

To examine the discriminatory metabolic phenotype between control and stroke risk patients, we compared the raw apLCMS table containing m/z values and intensities between control and stroke risk sera using PLS-DA. The apLCMS feature table containing 7197 features was inserted into SIMCA 14.1 (Umetrics AB, Umeå, Sweden). For increasing the accuracy of metabolite identification in data, set unit variance (UV) scaling was performed. As shown in Fig. 2A, the score plot of PLS-DA tends to separate the control subjects and stroke risk patients. An R^2 = 0.991 and Q^2 = 0.785, were observed in the permutation of the PLS-DA model, indicating that the quality of the computed model was not over-lifting and was reliable.[15] This result indicates that the serum metabolome of stroke risk patients was different from control subjects. Additionally, the Student's t-test performed in MetaboAnalyst 4.0, with a false discovery rate (FDR) adjusted q value of 0.05, showed that 2444 features (pink dots) were significantly different between control and stroke risk patients (Fig. 2A). The significant metabolites with q < 0.05 and first principal component of the

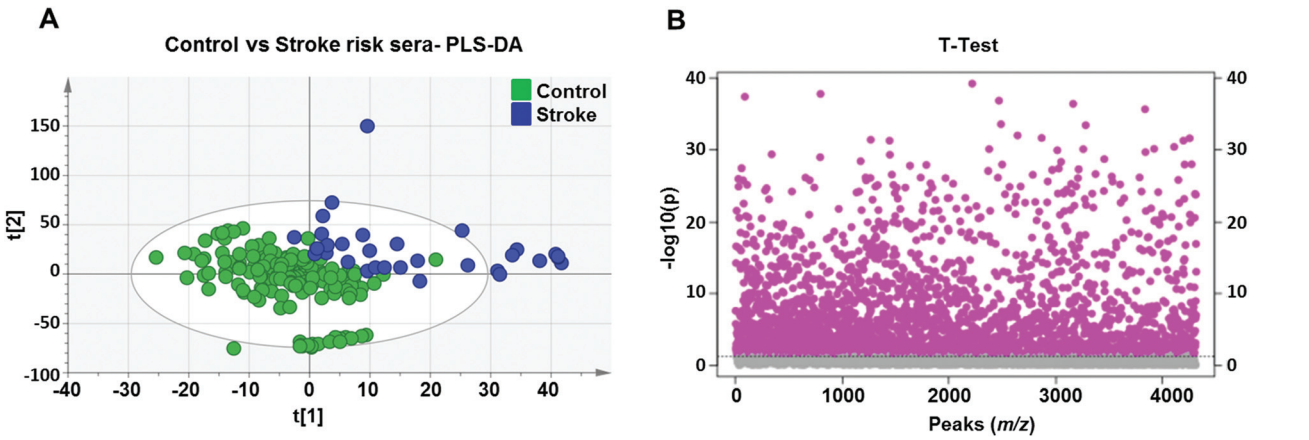

Fig. 2 Differential metabolic profiles of sera of stroke risk patients and control subjects. Separation and classification of the metabolic profile between stroke risk patient sera and control sera. (a) Discrimination of metabolome profile between stroke risk patient sera and control sera by two-dimensional partial least-squares discriminant analysis (PLS-DA), with R^2Y(cum) = 0.99 and Q^2(cum) = 0.793; six components. In the PLS-DA score plot, each data point represents one serum sample. Twenty permutations using six components resulted in R^2 = 0.991 and Q^2 = 0.785. (b) Important features selected by Manhattan plot using Student's t-test with a false discovery rate (FDR) adjusted p value threshold of 0.05 between stroke risk patient sera and control sera. The y-axis represents the −log 10 of the raw p value between compared groups, while the x-axis shows the compounds after normalization of their m/z values ranging from 50 to 1000. The dashed line represents the FDR significant threshold (q = 0.05), which separates the significant features as pink dots (2444 compounds out of 7197; p value <0.05) from other insignificant m/z values; black dots. Stroke represents stroke risk patients.

variable importance in projection (VIP) > 1, were considered to be influential for the separation of samples in PLS-DA analysis (Fig. 2).

3.3 Effect of hypertension, smoking, diabetes mellitus, and insulin sensitivity on metabolic phenotype of stroke risk patients

Risk factors of stroke such as HTN, smoking, DM, and IS were further taken into account while evaluating the specific differential metabolic phenomena caused by stroke occurrence. In order to exclude the possibilities of metabolic disturbance contamination by these risk factors, the above results were validated among control and stroke risk sera of HTN, smoking, DM, and IS subjects. As shown in Table 1, control subjects consisted of 79 hypertensive subjects, 82 smokers, 20 DM subjects, and 138 IS subjects. Similarly, among stroke risk patients, there were 51 hypertensive, 36 smokers, 20 DM, and 64 IS subjects. We presumed that the metabolic shift in the score plot of PLS-DA (Fig. 2) might be related to these risk factor. In order to exclude the possibility of distortion of data by the risk factors related to metabolic effect, we separately analyzed the hypertensive control with hypertensive stroke risk sera (Fig. 3A), smokers control with smokers stroke risk sera (Fig. 3B), DM control with DM stroke risk sera (Fig. 3C), and IS control with IS stroke risk sera. However, regardless of these risk factors among control and stroke groups, the score plot of PLS-DA efficiently separated the hypertensive control group from the hypertensive stroke risk group (Fig. 3A). Similarly, as shown in Fig. 3B, the smokers control group, DM control group (Fig. 3C), and IS control groups (Fig. 3D) were efficiently separated from smokers, DM, and IS stroke risk groups. These results suggest that the metabolic alterations among the control and stroke risk patients (Fig. 2) were specifically induced by the occurrence of thrombotic stroke and not HTN, smoking, DM, or IS.

3.4 Identification of the metabolic variations caused by stroke occurrence in patient sera

To identify differential metabolite signatures, we combined the VIP values generated from the PLS-DA model with the results from the two-tailed Student's *t*-test (Fig. 2). The KEGG database (http://www.kegg.jp), in combination with the results

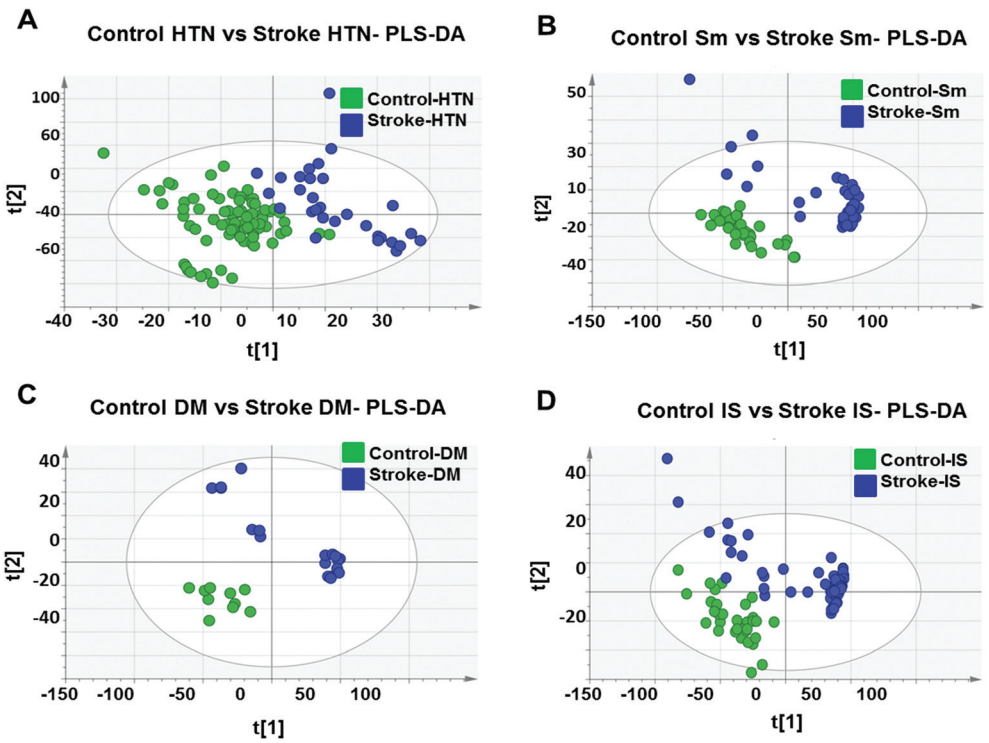

Fig. 3 Effect of hypertension, smoking, diabetes mellitus, and insulin sensitivity on metabolic profiles of stroke risk patients and control subject sera. (a) Discrimination of detected features between hypertensive stroke patient sera and hypertensive control sera by PLS-DA with R^2Y(cum) = 0.98 and Q^2(cum) = 0.89; five components. Twenty permutations using five components resulted in R^2 = 0.82 and Q^2 = −0.109. (b) Discrimination of detected features between smoker stroke risk patient sera and smoker control sera by PLS-DA with R^2Y(cum) = 0.99 and Q^2(cum) = 0.79; four components. Twenty permutations using four components resulted in R^2 = 0.988 and Q^2 = 0.639. (c) Discrimination of detected features between diabetic stroke risk patient sera and diabetic control sera by PLS-DA with R^2Y(cum) = 0.99 and Q^2(cum) = 0.67; three components. Twenty permutations using four components resulted in R^2 = 0.991 and Q^2 = 0.684. (d) Discrimination of detected features between insulin-sensitive stroke risk patient sera and insulin-sensitive control sera by PLS-DA with R^2Y(cum) = 0.99 and Q^2(cum) = 0.85; five components. Twenty permutations using five components resulted in R^2 = 0.99 and Q^2 = 0.704. In the PLS-DA score plot, each data point represents one serum sample. Variation described by each component of PLS-DA is given in the *x*- and *y*-axis labels. HTN represents hypertension, Sm represents smokers, DM represents diabetes mellitus, is represents insulin sensitivity.

from the analysis of METLIN (https://metlin.scripps.edu), was used to annotate the significant features obtained from the Student's *t*-test. As discussed earlier, the *t*-test yielded 2444 significant features between control and stroke risk sera. The recorder KEGG numbers of these metabolites were employed to identify the top affected pathways in the KEGG database. Considering the possible impact of affected pathways identified in KEGG, we measured the raw peak intensity of all the metabolites related to pathways. Raw intensity was measured by building bar graphs of each metabolite (data not shown). Biosynthesis of amino acid, purine metabolism, and lipid metabolism were considered to be influential for the separation of the distinct PLS-DA profile of stroke compared to control sera. As shown in Fig. 4 and ESI Fig. 1,† nineteen metabolites were selected based on an FC > 2 along with VIP = 1.5 and *q*-value <0.05: L-tryptophan (*m/z*: 205.09 $[M + H]^+$), 3-methoxytyramine (*m/z*: 150.09 $[M + H - H_2O]^+$), methionine (*m/z*: 150.05 $[M + H]^+$), *S*-adenosyl-L-methionine (*m/z*: 381.13 $[M + H - H_2O]^+$), cysteine (*m/z*: 241.03 $[M + H]^+$), homocysteinesulfinic acid (*m/z*: 168.03 $[M + H]^+$), leucine (*m/z*: 154.08 $[M + Na]^+$), isoleucine (*m/z*: 114.09 $[M + H - H_2O]^+$), proline (*m/z*: 98.06 $[M + H - H_2O]^+$), carnitine (*m/z*: 162.11 $[M + H]^+$), and arginine (*m/z*: 197.10 $[M + Na]^+$) in the biosynthesis of amino acid pathway; dopamine (*m/z*: 154.08 $[M + H]^+$) in the neurotransmitter pathway; 2-oxoglutarate (*m/z*: 169.01 $[M + Na]^+$) and NAD^+ (*m/z*: 686.09 $[M + Na]^+$) in the TCA cycle; guanosine (*m/z*: 284.09 $[M + H]^+$) in purine metabolism; and linoleic acid (*m/z*: 281.24 $[M + H]^+$), sphingosine (*m/z*: 300.28 $[M + H]^+$), 3-ketosphingosine (*m/z*: 280.26 $[M + H - H_2O]^+$), and cerebrosterol (*m/z*: 403.35 $[M + H]^+$) in the lipid metabolism pathway. These metabolites were significantly elevated among the stroke risk sera and were related to top affected pathways. Sixteen metabolites: phenylalanine (*m/z*: 166.08 $[M + H]^+$), valine (*m/z*: 140.06 $[M + Na]^+$), L-homocysteine (*m/z*: 136.04 $[M + H]^+$), cysteine sulfinic acid (*m/z*: 136.006 $[M + H - H_2O]^+$), lysine (*m/z*: 185.06 $[M + K]^+$), and *N*-acetylaspartate (*m/z*: 214.01 $[M + K]^+$) in biosynthesis of the amino acid pathway; serotonin (*m/z*: 194.12 $[M + NH_4]^+$), *N*-acetylserotonin (*m/z*: 201.10 $[M + H - H_2O]^+$), and epinephrine (*m/z*: 166.08 $[M + H - H_2O]^+$) in the neurotransmitter pathway; and ATP (*m/z*: 529.98 $[M + Na]^+$), AMP (*m/z*: 370.05 $[M + Na]^+$), GTP (*m/z*: 523.99 $[M + H]^+$), ITP (*m/z*: 490.97 $[M + H - H_2O]^+$), dATP (*m/z*: 513.98 $[M + Na]^+$), and dAMP (*m/z*: 513.98 $[M + H - H_2O]^+$) in purine metabolism were significantly downregulated in the stroke risk sera.

3.5 Validation and quantification of selected metabolites in the stroke sera by MS/MS

Collective detection of the abovementioned untargeted 35 serum metabolites in sera will have the most power for diagnosis of stroke risk. However, diagnosis based on quantification of many biomarkers is not convenient or economical in clinical practice. To identify simplified serum metabolite signatures that would be more practical for use in diagnosing stroke patients, the subset of 35 putative metabolites was tested by MS/MS to validate our previous result (Fig. 4 and ESI Fig. 1†). The presence of 16 key metabolites: *N*-acetylserotonin, serotonin, L-tryptophan, 3-methoxytyramine, methionine, homocysteinesulfinic acid, cysteine, cysteine sulfinic acid, isoleucine, valine, lysine, carnitine, arginine, *N*-acetylaspartate, linoleic acid, and sphingosine were confirmed in stroke risk sera by comparing their spectra with the standards available in the HMDB databases (http://www.hmdb.ca), as well as with the MS/MS spectra of the standard chemicals. These compounds were scanned followed by product-ion analysis using the col-

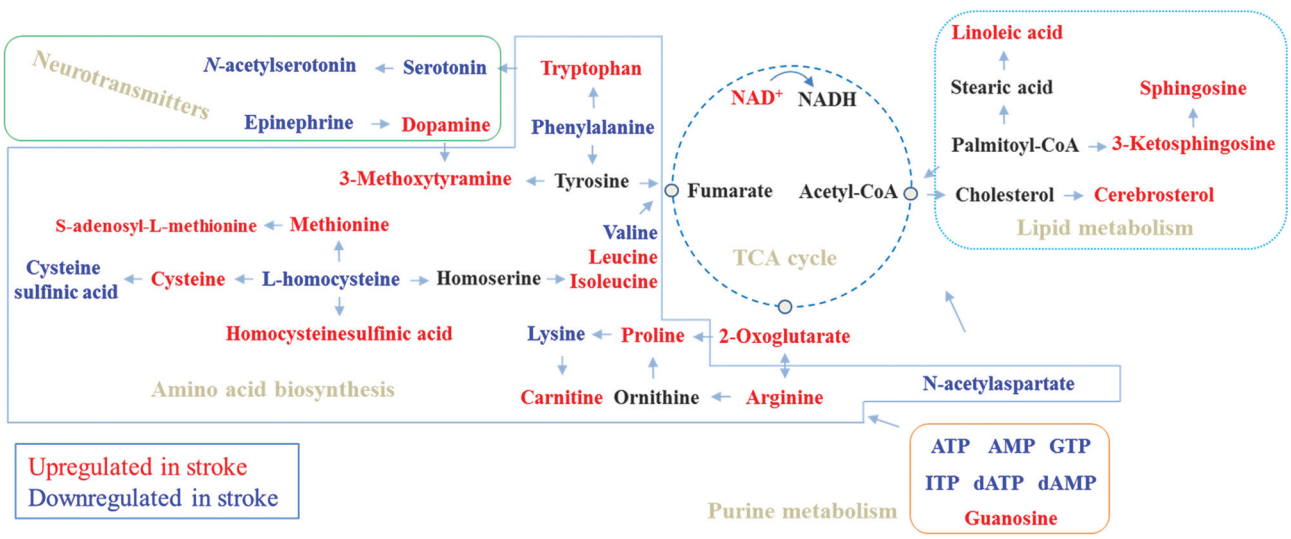

Fig. 4 Schematic overview of the disturbed metabolic pathways and associated metabolites among stroke risk patients. The red-colored metabolite names represent the metabolites with significantly high intensity among stroke risk sera compared with control sera. The blue-colored metabolite names represent the metabolites with significantly low intensity among stroke risk sera compared with control sera. The metabolites in black font are those that were not found in the sera of either stroke or control patients but were interconnected to our identified metabolites. The brown-colored fonts are the pathway names for the identified interconnected (boxed) metabolites.

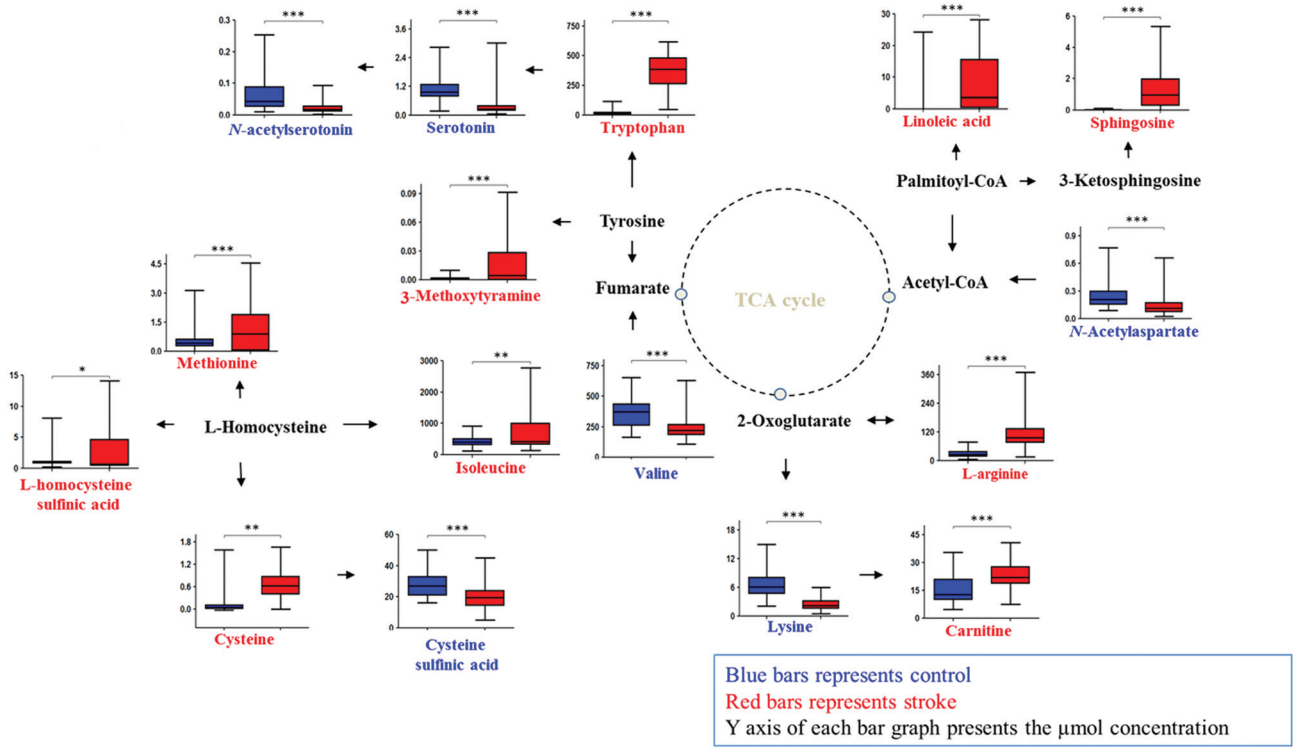

Fig. 5 Pathway overview and quantified concentration of metabolic signatures in serum samples by LC-ESI/MS/MS. (a) Quantified (micro molar) concentrations of L-tryptophan, 3-methoxytyramine, methionine, homocysteinesulfinic acid, cysteine, isoleucine, carnitine, arginine, linoleic acid, sphingosine, *N*-acetylserotonin, serotonin, cysteine sulfinic acid, valine, lysine, and *N*-acetylaspartate serum samples from stroke risk patients and control subjects. Quantified concentration was determined in reference to the calibration curve of each standard compound. Concentrations of each compound were calculated by reference to the peak areas of the external standards within the range of LOD and LOQ. Stroke represents stroke risk patients. ***$p < 0.001$; **$p < 0.01$ by Student's *t*-test.

lision energy values 0, 5, 10, 15, and 20 eV. LC-MS/MS scan mode, product-ion mode, and multiple reaction monitoring (MRM) parameters of the $[M + H]^+$ ion for the quantified compounds in stroke risk sera are shown in ESI Table 1.† The concentrations of these compounds were determined in control and stroke sera, and the results are shown in Fig. 5. Their concentrations in serum were calculated by referring to the external standard's calibration curve. In accordance with our LC/MS results (Fig. 4 and ESI Fig. 1†), the quantified concentrations of L-tryptophan, 3-methoxytyramine, methionine, homocysteinesulfinic acid, cysteine, isoleucine, carnitine, arginine, linoleic acid, and sphingosine were elevated in stroke risk sera, while the quantified concentrations of *N*-acetylserotonin, serotonin, cysteine sulfinic acid, valine, lysine, and *N*-acetylaspartate were down-regulated in stroke risk sera (Fig. 5).

3.6 Effect of hypertension, smoking, diabetes, and insulin sensitivity on quantified biomarkers of stroke

To exclude the possibilities of hypertension-, smoking-, diabetes-, and insulin sensitivity-related effects on the validated and quantified concentrations of the potential metabolites given in Fig. 5, we measured the concentrations of 16 key metabolites: *N*-acetylserotonin, serotonin, L-tryptophan, 3-methoxytyramine, methionine, homocysteinesulfinic acid, cysteine, cysteine sulfinic acid, isoleucine, valine, lysine, carnitine, arginine, *N*-acetylaspartate, linoleic acid, and sphingosine in the control group with HTN, Sm, DM and IS against patients at risk for stroke with HTN, Sm, DM and IS. As pointed out by red arrows (Fig. 6), in contrast to Fig. 5 the quantified concentrations of *N*-acetylserotonin were non-significant in the hypertensive control and hypertensive stroke risk sera (Fig. 6). Similarly, valine and cysteine sulfinic acid were non-significant in diabetic control and diabetic stroke risk sera (Fig. 6). While, isoleucine showed no significantly different concentration by the presence of smoking, diabetes and insulin sensitivity among the control and stroke risk groups (Fig. 6). However, in accordance with Fig. 5, all putative metabolites showed a similar trend of up- or down-regulation in stroke risk sera (Fig. 6). Interestingly, thirteen metabolites among the selected 16 metabolite's panel showed significantly different concentrations between the control group with HTN, Sm, DM and IS and patients at risk for stroke with HTN, Sm, DM and IS which were in accordance with Fig. 5. This result further provides evidence that the metabolites alterations reported in this study were specifically induced by the occurrence of stroke and were not related to the risk factors.

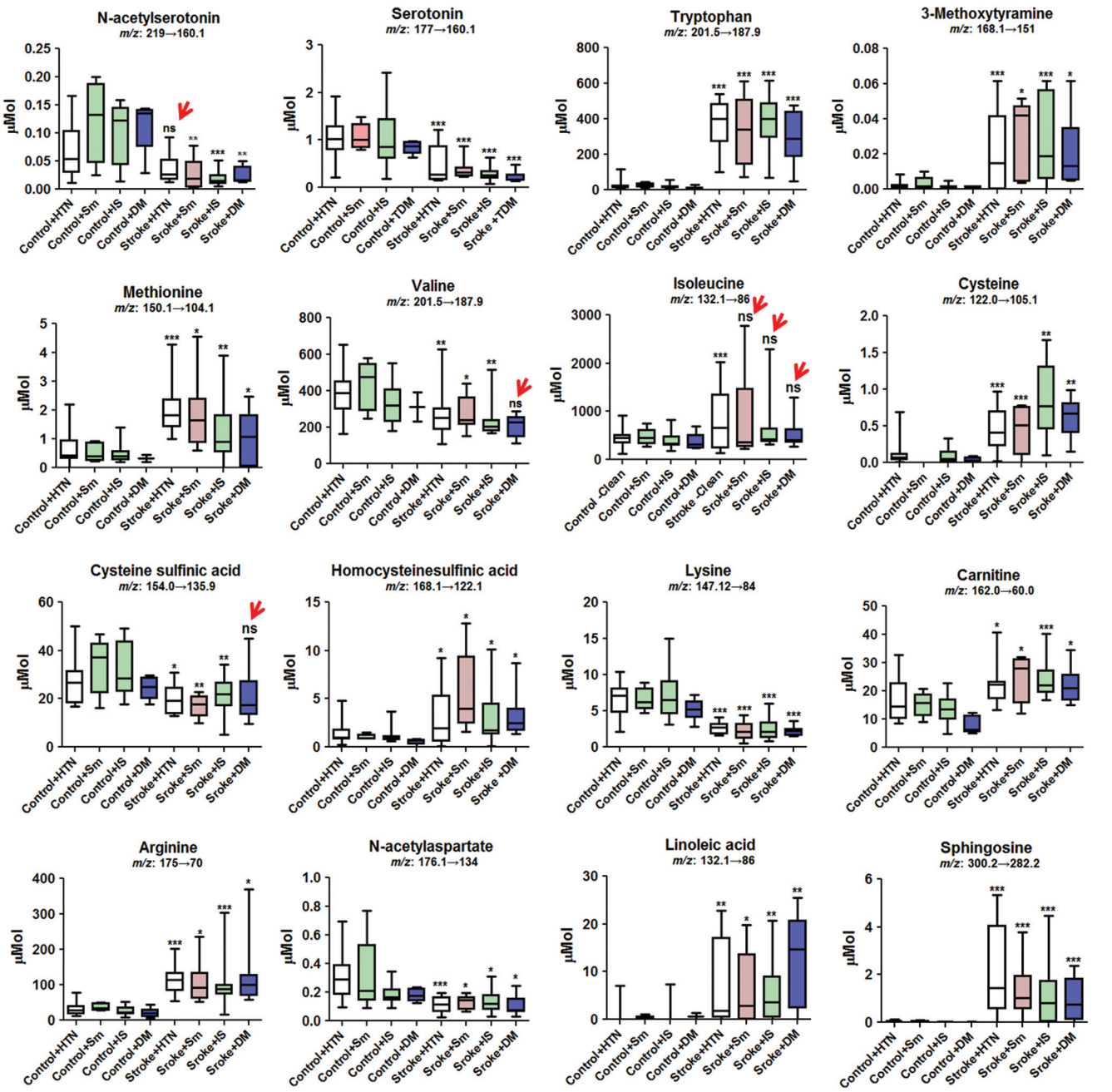

Fig. 6 Effect of hypertension, smoking, diabetes and insulin sensitivity on quantified concentrations of metabolic signatures in serum samples by LC-ESI/MS/MS. Quantified (micromolar) concentrations of *N*-acetylserotonin, serotonin, L-tryptophan, 3-methoxytyramine, methionine, homocysteinesulfinic acid, cysteine, cysteine sulfinic acid, isoleucine, valine, lysine, carnitine, arginine, *N*-acetylaspartate, linoleic acid, and sphingosine in the control group with HTN (n = 28), Sm (n = 08), DM (n = 08) and is (n = 15) against patients at risk for stroke with HTN (n = 12), Sm (n = 09), DM (n = 08) and is (n = 21). HTN represents hypertension, Sm represents smokers, DM represents diabetes mellitus, is represents insulin sensitivity. Quantified concentration was determined in reference to the calibration curve of each standard compound. Concentrations of each compound were calculated by reference to the peak areas of the external standards within the range of LOD and LOQ. Stroke represents stroke risk patients. ***$p < 0.001$; **$p < 0.01$ by Student's t-test.

4. Discussion

Previously, we showed decreased levels of the catabolites lysine and valine in the sera of patients with thrombotic risk.[3] Additionally, elevated homocysteinsulfinic acid and oxoglutarate along with deceased levels of ubiquinone further supported the low levels of lysine and valine catabolites.[3] Here, we further extended our research in order to observe the global overview of metabolic disturbance in a separate set of people who were later diagnosed with thrombotic stroke. In this study, we employed both untargeted and targeted metabolomics in order to validate our previous results as well as to determine

the global disturbances in the metabolome of thrombotic stroke risk sera. This study identified a set of 35 putative metabolites (19 upregulated and 16 downregulated in stroke risk sera) using untargeted metabolomics. Further validation and quantification of these 35 metabolites using targeted metabolomics confirmed the presence of 16 key metabolites (10 upregulated and 6 downregulated) in stroke risk sera.

Over the last decade, metabolome profiling has been applied in stroke in several studies using animal models and stroke patients.[15] Especially, PLS-DA gained popularity in the field of metabolomics analysis for distinguishing stroke patients from controls.[17] However, in this study, we employed metabolome profiling techniques on people who were not diagnosed with stroke yet. Here, in accordance with our previous study, we observed that the score plot obtained from PLS-DA was able to distinguish the patients who were not yet diagnosed with stroke from healthy subjects, indicating that there was a different serum metabolome in the patients who were developing stroke disease.[3] Further, using PLS-DA, we observed that these metabolic alterations between stroke risk subjects and control were specifically related to stroke occurrence and were not related to other factors such as HTN, smoking, DM, and IS. Additionally, a majority of the previous studies employed untargeted metabolomics.[5] With few exceptions, the results of many studies are non-reproducible because of the lack of targeted metabolomics.[3,5] However, we used both the untargeted and targeted metabolomics profiling. Using untargeted metabolomics, we detected a set of putative metabolic end-products, and while with the targeted metabolomics we were able to further confirm a set of putative biomarkers that further supported the specificity, accuracy, precision, and stability of our untargeted metabolomics results.

Many diabetic patients are at risk for stroke and peripheral vascular disease, and smoking is associated with worsening of metabolic control in diabetic patients.[3,18] In addition, HTN, which may lead to increased stress on the endothelium, can cause endothelial damage and an altered blood cell–endothelium interaction, which can lead to local thrombi formation and ischemic lesions.[19] Nonetheless, HTN, smoking, and DM can also cause a multitude of metabolic changes in the serum metabolome. Hence, in this study, we considered smoking, HTN, and DM along with IS as risk factors to evaluate the metabolic perturbation that may have been specifically caused by stroke. However, whether the stoke occurrence was due to the effect of smoking or a surrogate of the HTN and DM, our results, based on PLS-DA, found negligible effects of these factors in the metabolic alterations among stroke risk patients.

There is a growing body of evidence affirming the association between stroke outcomes and glutamate-induced excitotoxity, possibly because of the oxidative stress through activation of nitric oxide synthase, inflammation through expression of inflammatory cytokines, and endothelial damage through matrix metalloproteinase (MMP) release.[20] In our previous study, we detected lysine as an important precursor of glutamate along with elevated level of L-homocysteine sulfinic acid (L-HSA) and a decrease in ubiquinone levels, which may have synergized to initiate excitotoxity in stroke risk patients.[3] Consistent with our previous study, in this study, our targeted metabolome profiling showed similar decreased levels of lysine and valine in patients with stroke risk. In addition, this study further confirmed the previously reported elevated level of L-HSA in stroke risk sera.[3,20] Elevated L-HSA, along with carnitine, tryptophan, 3-methoxytyramine, and sphingosine (Fig. 7), are associated with stroke severity and worse neuro-

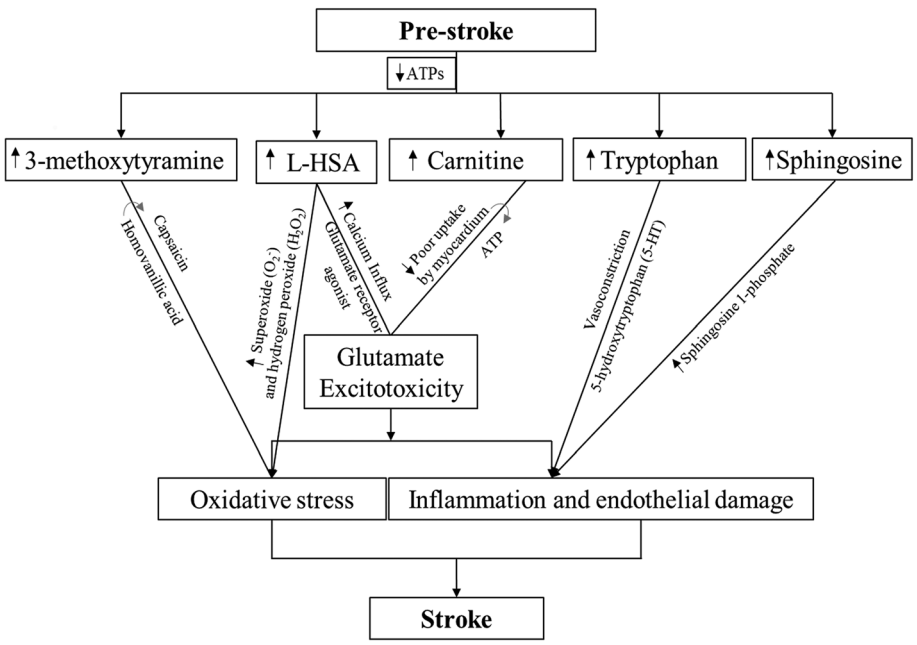

Fig. 7 Schematic diagram of identified interrelated subsets of early biomarkers in stroke risk subjects their impact on stroke events.

logical outcomes in patients with stoke risk.[21] On the basis of these identified metabolites and their synergistic impact on stroke events, our study proposes new mechanistic insight into the development of events leading to thrombotic stroke.

Most of the identified and validated metabolites discovered in this study fall into three pathophysiological categories of developing stroke: glutamate excitotoxicity, oxidative stress due to formation of free radicals, and cell-mediated inflammatory response and endothelial damage.[22] A high serum level of L-HSA arose as a potential predictor of thrombotic stroke in our current as well as previous study.[3] In addition, we detected a high level of L-HSA in patients at risk for acute myocardial infarction (unpublished data). Such an elevated level of L-HSA is known to exert excitotoxicity on *N*-methyl-D-aspartate (NMDA) receptors *via* exhibiting L-glutamic acid- and L-aspartic acid-like effects.[23] Elevated glutamate can trigger excitotoxity, which leads to oxidative stress *via* activation of nitric oxide synthase, inflammation *via* the expression of inflammatory cytokines, and endothelial damage *via* MMP release.[24] In this study, we could not detect glutamate in thrombotic stroke risk sera, which might be due to a shorter duration of increase in glutamate after the onset of ischemic stroke.[25] Hence, this may have not been detected because the serum was obtained before stroke occurrence. Nonetheless, elevated carnitine along with L-HSA supports our hypothesis for glutamate excitotoxicity, since high carnitine is required for transporting long-chain fatty acids to produce energy in the mitochondrial matrix, a main source of ATP production in ischemic cells.[26] The exact mechanism for its increase is unknown, but we hypothesize that the elevated serum level of carnitine might have resulted from the poor uptake by the ischemic brain.[27]

Interestingly, our untargeted metabolomics results also showed downregulated ATPs in stroke risk sera along with other metabolites involved in purine metabolism. Such lowered ATP production and L-HSA, as a glutamate receptor agonist, may modulate activities of mGluRs, predisposing individuals to the neurotoxic consequences of stroke.[28] In addition, the generation of superoxide anion radical and hydrogen peroxide (H_2O_2) with elevation of L-HAS,[29] along with 3-methoxytyramine-related capsaicin-induced generation of intracellular reactive oxygen species (ROS) can cause oxidative injuries in cerebral ischemia.[30] This result was further supported by lowered *N*-acetylserotonin (NAS) and serotonin levels in stroke risk sera, since NAS possesses known properties of neuroprotection by inhibiting mitochondrial death pathways[31] and was reported to play a role in protection against oxidative stress injuries caused by H_2O_2.[32–34] Similarly, serotonin exerts powerful antioxidant actions;[35] however, its degradation is shown to generate ROS through the mitochondrial enzyme monoamine oxidase A (MAO-A).[36] However, contrary to the key mechanisms described in animal studies, we did not investigate the ROS and H_2O_2 production in ischemic sera, underlining the importance of validating these results in stroke patient sera in future studies.

Our study also highlights the disruptions in tryptophan concentration in stroke risk patient sera possibly due to the conversion of tryptophan to 5-hydroxytryptophan (5-HT),[37,38] which may contribute to vasoconstriction[38,39] and inflammatory response corresponding to endothelial damage.[40] In accordance, consistent with the previous stroke studies in humans and animals,[1,41] sphingolipid was significantly increased in stroke risk sera, suggesting that changes in the sphingolipid profile arose during the developmental stages of stroke. Moreover, we detected a disturbance in the interconnected metabolites L-cysteine sulfinic acid, L-cysteine, and their precursor, L-methionine. L-Cysteine is produced through cleavage of cystathionine by cystathionine beta synthase. Through validating and quantifying potential metabolites by MS/MS, we detected upregulated L-cysteine and L-methionine in stroke risk sera and downregulated L-cysteine sulfinic acid; however, the exact mechanism is still unknown.

Despite these findings, our study has some limitations. First, our study lacks the availability of a replication cohort and therefore, validations should be carried out on a larger population. Second, due to insufficient availability of serum samples, a small number of sera samples were quantified to investigate the effect of hypertension, smoking, diabetes, and insulin sensitivity on quantified biomarkers of stroke. Therefore validation and quantification of the selected metabolite panel needs to be carried out on a larger population of control and stroke risk subjects to investigate the risk factor's related metabolic alterations.

5. Conclusions

Our study provides valuable insight into the global view of the metabolic perturbations associated with three pathophysiological categories of developing stroke: glutamate excitotoxicity, oxidative stress due to formation of free radicals, and cell-mediated inflammatory response and endothelial damage. In support of our previous study, elevated homocysteine sulfinic acid, along with the biomarkers identified herein of methionine, cysteine, 3-methoxytyramine, carnitine, tryptophan, sphingosine, linoleic acid, arginine, and isoleucine, were observed as a panel of potential biomarkers for ischemic stroke. They can be used for early stage disease risk identification, early diagnosis, pathological mechanism research, and drug target screening. Further studies are required to validate and generalize the applicability of these potential metabolites as novel biomarkers in the management of ischemic stroke patients.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank Korea Health Industry Development Institute (grant no. HI14C2686), National Research Foundation of Korea (grant no. NRF-2017R1A2B4003890 &

NRF-2017M3A9F1031229), and Cooperative Research Program for Agriculture Science and Technology Development [Project No. PJ01345402] Rural Development Administration, Republic of Korea, for financially supporting this study. The authors thank Dr Karan Uppal and Dr Shuzhao Li from the Emory University School of Medicine (Atlanta, GA, USA) for providing the R-package used to run the apLCMS and xMSanotator.

References

1 N. Sun, R. F. Keep, Y. Hua and G. Xi, *Transl. Stroke Res.*, 2016, **7**, 420–438.
2 C. J. Murray, *et al.*, *Lancet*, 2015, **386**, 2145–2191.
3 Y. Lee, A. Khan, S. Hong, S. H. Jee and Y. H. Park, *Mol. BioSyst.*, 2017, **13**, 1109–1120.
4 M. A. Moskowitz, E. H. Lo and C. Iadecola, *Neuron*, 2010, **67**, 181–198.
5 S. J. Kim, G. J. Moon and O. Y. Bang, *J. Stroke*, 2013, **15**, 27–37.
6 E. E. Smith, G. Saposnik, G. J. Biessels, F. N. Doubal, M. Fornage, P. B. Gorelick, S. M. Greenberg, R. T. Higashida, S. E. Kasner and S. Seshadri, *Stroke*, 2017, **48**, e44–e71.
7 M. C. Leary and J. L. Saver, *Cerebrovasc. Dis.*, 2003, **16**, 280–285.
8 A. Lopez-Lopez, A. Lopez-Gonzalvez, T. C. Barker-Tejeda and C. Barbas, *Expert Rev. Mol. Diagn.*, 2018, **18**, 557–575.
9 D. K. Trivedi, K. A. Hollywood and R. Goodacre, *New Horiz. Transl. Med.*, 2017, **3**, 294–305.
10 Y. H. Jee, J. Emberson, K. J. Jung, S. J. Lee, S. Lee, J. H. Back, S. Hong, H. Kimm, P. Sherliker, S. H. Jee and S. Lewington, *Int. J. Epidemiol.*, 2018, **47**, 385–386f.
11 A. Khan, O. S. Shin, J. Na, J. K. Kim, R.-K. Seong, M.-S. Park, J. Y. Noh, J. Y. Song, H. J. Cheong, Y. H. Park and W. J. Kim, *Sci. Rep.*, 2019, **9**, 4760.
12 A. Khan, S. A. Choi, J. Na, A. D. Pamungkas, K. J. Jung, S. H. Jee and Y. H. Park, *J. Proteome Res.*, 2019, **18**, 1532–1541.
13 A. Khan, H. Park, H. A. Lee, B. Park, H. S. Gwak, H.-R. Lee, S. H. Jee and Y. H. Park, *Toxicol. Sci.*, 2017, **160**, 371–385.
14 T. Yu, Y. Park, J. M. Johnson and D. P. Jones, *Bioinformatics*, 2009, **25**, 1930–1936.
15 M. N. Triba, L. Le Moyec, R. Amathieu, C. Goossens, N. Bouchemal, P. Nahon, D. N. Rutledge and P. Savarin, *Mol. BioSyst.*, 2015, **11**, 13–19.
16 K. Uppal, D. I. Walker and D. P. Jones, *Anal. Chem.*, 2017, **89**, 1063–1067.
17 W. H. Heijne, R. J. Lamers, P. J. van Bladeren, J. P. Groten, J. H. van Nesselrooij and B. van Ommen, *Toxicol. Pathol.*, 2005, **33**, 425–433.
18 B. Eliasson, *Prog. Cardiovasc. Dis.*, 2003, **45**, 405–413.
19 B. B. Johansson, *Clin. Exp. Pharmacol. Physiol.*, 1999, **26**, 563–565.
20 M. Castellanos and J. Serena, *Cerebrovasc. Dis.*, 2007, **24**(Suppl 1), 7–15.
21 C. J. Chang, H. Ishii, H. Yamamoto, T. Yamamoto and M. Spatz, *J. Neurochem.*, 1993, **60**, 1483–1490.
22 E. Sidorov, D. K. Sanghera and J. K. P. Vanamala, *J. Stroke*, 2019, **21**, 31–41.
23 C. R. Santhosh-Kumar, J. C. Deutsch, J. C. Kolhouse, K. L. Hassell and J. F. Kolhouse, *Anal. Biochem.*, 1994, **220**, 249–256.
24 D. Song, M. H. O'Regan and J. W. Phillis, *Neurosci. Lett.*, 1996, **220**, 1–4.
25 A. Dávalos, J. Castillo, J. Serena and M. Noya, *Stroke*, 1997, **28**, 708–710.
26 S. S. Moselhy and S. H. Demerdash, *Clin. Biochem.*, 2009, **42**, 78–82.
27 M. E. Pierpont, D. Judd, I. F. Goldenberg, W. S. Ring, M. T. Olivari and G. L. Pierpont, *Am. J. Cardiol.*, 1989, **64**, 56–60.
28 S. Ramakrishnan, K. N. Sulochana, S. Lakshmi, R. Selvi and N. Angayarkanni, *Indian J. Biochem. Biophys.*, 2006, **43**, 275–283.
29 J. Loscalzo, *J. Clin. Invest.*, 1996, **98**, 5–7.
30 K. Ito, T. Nakazato, K. Yamato, Y. Miyakawa, T. Yamada, N. Hozumi, K. Segawa, Y. Ikeda and M. Kizaki, *Cancer Res.*, 2004, **64**, 1071–1078.
31 H. Zhou, J. Wang, J. Jiang, I. G. Stavrovskaya, M. Li, W. Li, Q. Wu, X. Zhang, C. Luo, S. Zhou, A. C. Sirianni, S. Sarkar, B. S. Kristal, R. M. Friedlander and X. Wang, *J. Neurosci.*, 2014, **34**, 2967–2978.
32 J. Jiang, S. Yu, Z. Jiang, C. Liang, W. Yu, J. Li, X. Du, H. Wang, X. Gao and X. Wang, *Oxid. Med. Cell. Longevity*, 2014, **2014**, 310504.
33 H. Zhou, J. Wang, J. Jiang, I. G. Stavrovskaya, M. Li, W. Li, Q. Wu, X. Zhang, C. Luo, S. Zhou, A. C. Sirianni, S. Sarkar, B. S. Kristal, R. M. Friedlander and X. Wang, *J. Neurosci.*, 2014, **34**, 2967–2978.
34 J.-M. Yoo, B. D. Lee, D.-E. Sok, J. Y. Ma and M. R. Kim, *Redox Biol.*, 2017, **11**, 592–599.
35 J. R. Munoz-Castaneda, P. Montilla, F. J. Padillo, I. Bujalance, M. C. Munoz, J. Muntane and I. Tunez, *Acta Neurobiol. Exp.*, 2006, **66**, 1–6.
36 A. Nocito, F. Dahm, W. Jochum, J. H. Jang, P. Georgiev, M. Bader, E. L. Renner and P. A. Clavien, *Gastroenterology*, 2007, **133**, 608–618.
37 A. Grafka, M. Łopucki, K. K-. Kajszczarek, M. S-. Kosarzycka and G. Dzida, *Plasma concentration of tryptophan and pregnancy-induced hypertension*, 2018.
38 S. W. Watts, S. F. Morrison, R. P. Davis and S. M. Barman, *Pharmacol. Rev.*, 2012, **64**, 359–388.
39 G. S. Allen, L. M. Henderson, S. N. Chou and L. A. French, *J. Neurosurg.*, 1974, **40**, 442–450.
40 M. S. Shajib and W. I. Khan, *Acta Physiol.*, 2015, **213**, 561–574.
41 S. A. Sheth, A. T. Iavarone, D. S. Liebeskind, S. J. Won and R. A. Swanson, *PLoS One*, 2015, **10**, e0129735.