Circulating metabolites and risk of stroke in seven population-based cohorts

Dina Vojinovic, Marita Kalaoja, Stella Trompet, Krista Fischer, Martin J. Shipley, Shuo Li, Aki S. Havulinna, Markus Perola, Veikko Salomaa, Qiong Yang, Naveed Sattar, Pekka Jousilahti, Najaf Amin, Ramachandran S. Vasan, M. Arfan Ikram, Mika Ala-Korpela, J. Wouter Jukema, Sudha Seshadri, Johannes Kettunen, Mika Kivimaki, Tonu Esko, Cornelia M. van Duijn

This chapter is in preparation.

The supplemental information for this paper is available at https://drive.google.com/file/d/1XXBT5NWkjSvq_MDHaWDlh_mXe6MqtqgV/view?usp=sharing

ABSTRACT

Stroke is a leading cause of death and long-term disability worldwide. Detailed profiling of metabolic status can provide insights into metabolic changes and lead to identification of individuals with higher risk of stroke. We investigated association of a wide range of metabolites with risk of stroke in seven prospective population-based cohorts including 1,791 incident stroke events among 38,797 participants. The analyses were performed considering all incident stroke events and ischemic and hemorrhagic events separately. The analysis revealed ten significant metabolite associations. Amino acid histidine (hazard ratio (HR) per SD= 0.9, 95% confidence interval (CI): 0.85, 0.94), glycolysis-related metabolite pyruvate (HR per SD = 1.09, 95% CI: 1.04, 1.14), acute phase reaction marker glycoprotein acetyls (HR per SD = 1.09, 95% CI: 1.03, 1.15), cholesterol in high-density lipoprotein (HDL) 2 and several other lipoprotein particles including cholesterol in medium HDL and triglycerides in medium and large low-density lipoprotein (LDL) particles were associated with risk of stroke. When focusing on incident ischemic stroke, a significant association was observed with phenylalanine (HR per SD = 1.12, 95% CI: 1.05, 1.19) and total and free cholesterol in large HDL particles. When comparing our findings to those of a study in the China Kadoorie Biobank, glycoprotein acetyls are replicated both in Caucasians and Chinese. However, we also observed very significant associations that were specific for Western societies. To conclude, we found association of amino acids, glycolysis-related metabolites, acute phase reaction markers, and several lipoprotein subfractions with the risk of stroke. The biological mechanisms underlying these associations should be subject of further studies.

INTRODUCTION

Stroke is a leading cause of death and serious long-term disability worldwide.[1] The majority of strokes are of the ischemic type, while the hemorrhagic type occurs less often but is associated with a higher mortality risk.[1,2] Stroke risk is determined by various modifiable risk factors such as hypertension, diabetes mellitus, cardiovascular disease, smoking, and obesity, whereas association of stroke with cholesterol and its subfractions has shown inconsistent results.[1-6] Opportunities for therapeutic interventions in stroke patients depend on the type of stroke and relay on brain imaging techniques.[7] Despite advances in brain imaging techniques, costs are still high, availability is limited and not all patients show a relevant lesion on neuroimaging.[7,8] New technology is needed to identify high-risk patients, to understand the etiology of stroke and develop future prevention strategies. Detailed profiling of metabolic status can provide insights into metabolic changes that lead to a higher risk of stroke. As the metabolome reflects both genome and exposome including exposures to risk factors that determine the risk of stroke, this new –omics technology may open new avenues towards stroke prevention. To date, only few studies have analyzed metabolic disturbances in stroke and identified various metabolites to be associated with stroke.[9-11] However, these studies are based on a relatively small sample or on participants of non-European ancestry.[12] The most comprehensive study to date was conducted by Holmes *et al.* within the China Kadoorie Biobank including patients with ischemic stroke (N = 1,146) and intracerebral hemorrhage (N = 1,138).[12] The study reported an association between lipids and lipoprotein particles of various sizes with ischemic stroke but not with hemorrhage.[12] Furthermore, the study identified glycoprotein acetyls, ketone bodies, glucose and docosahexaenoic acid to be associated with both ischemic and hemorrhagic stroke.[12]

As large metabolomic studies of stroke in persons of European origin are lacking, the aim of our study is to conduct a comprehensive analysis of circulating metabolites and incident stroke in large prospective population-based setting including 1,791 incident stroke events among 38,797 participants of European origin.

METHODS

Study population

Our study population included 38,797 participants from seven cohorts including Rotterdam Study, the Whitehall II study (Whitehall II), the national FINRISK studies (FINRISK97 and FINRISK07), PROspective Study of Pravastatin in the Elderly at Risk (PROSPER), Estonian biobank (EGCUT), and Framingham Heart Study (FHS). Description of participating

studies is provided in the **Supplementary Note 1**. Each of the participating studies was approved by local ethical committees or institutional review boards. All participants provided written informed consent.

Stroke assessment

Details on stroke assessment are provided in the **Supplementary Note 2.** The incident stroke events were assessed through follow-up of health records, while in some studies additional periodic visits to research centers were used (e.g. Rotterdam Study, FHS). Participants of the Rotterdam Study were monitored for incident stroke using automated linkage of medical records from general practitioners with the study database.[13] Incident stroke events in the Whitehall II study were ascertained through linkage to electronic records from hospitalizations due to stroke and national statistics death registries,[14,15] whereas in the FINRISK cohorts linkage to national health registries was used (https://www.biorxiv.org/content/early/2018/03/12/280677). Ascertainment of incident stroke events in EGCUT was also performed through linkage to electronic records from multiple databases (https://thl.fi/publications/morgam/cohorts/ full/estonia/est-esta.htm), while information regarding domiciliary visits or hospitalizations associated with possible cardiovascular events including stroke, and information on all deaths was used for classification of study endpoints in PROSPER.[16] In the FHS, incident clinical stroke was identified as part of ongoing clinic and hospital surveillance, and additional stroke surveillance by annual phone health updates and collaboration with primary care physicians and local emergency departments.[17,18] Participants with a history of stroke at baseline were excluded from the analyses.

Other measurements

The baseline measurements included measures of blood pressure, plasma glucose levels, weight, and height. Hypertension was defined as a systolic blood pressure ≥ 140 mmHg, diastolic blood pressure ≥ 90 mmHg, or use of antihypertensive medication. Diabetes was defined as fasting plasma glucose levels above 7 mmol/L or use of medication indicated for the treatment of diabetes. Body mass index (BMI) was calculated as weight in kilograms divided by square of heights in meters.

Metabolite quantification

Circulating metabolites were quantified using a high-throughput Nuclear Magnetic Resonance (^{1}H-NMR) technology. In all participating studies except the FHS, the Nightingale Health metabolomics platform (Helsinki, Finland) was used for simultaneous quantification of a wide range of metabolites, including routine lipids, 14 lipoprotein subclasses and their lipids (esterified cholesterol, free cholesterol, total cholesterol, triglycerides, phospholipids, and total lipids), fatty acids, amino acids, ketone bodies,

and various glycolysis precursors. A detailed description of the methodology has been provided previously.[19,20] In the FHS, lipoprotein subclasses were measured by proton NMR spectroscopic assay (LipoScience, Raleigh, NC).[21,22] Blood samples were collected after overnight fasting in all studied except for FINRISK97 and FINRISK07 in which the samples were collected after 4 hours of fasting (semi-fasting state).[23,24] The sample material was EDTA-plasma in the Rotterdam Study, FHS, and EGCUT, whereas the serum was used in FINRISK, PROSPER and Whitehall II.[23-26] There were 147 primary non-derived metabolite measurements quantified in absolute concentration units that were further analyzed in this study.

Statistical analyses

To obtain an approximately normal distribution, all metabolites were natural logarithmic transformed prior the analyses. To deal with zero values, one was added to all values of the metabolites prior to the transformation. The metabolite measurements were subsequently scaled to standard deviation (SD) units. The relationship between metabolites and stroke was assessed by Cox proportional hazards regression models. The analyses were performed while adjusting for age, gender, BMI, lipid-lowering medication, and study-specific covariates if needed (Model 1). The associations were further adjusted for smoking status, diabetes, and hypertension (Model 2). The summary statistics results of participating studies were combined using inverse variance-weighted fixed-effect meta-analysis. The analyses were performed considering all incident stroke events and ischemic and hemorrhagic events separately.

As metabolite measures are highly correlated, we calculated the number of independent tests using the previously described method of Li and Ji.[27] Subsequently, the number of independent tests was used for calculation of Bonferroni corrected *p*-value (0.05/30 independent metabolites = 1.7×10^{-3}).

To determine the discrimination power of metabolite measures discovered in our study, we used the Rotterdam Study to calculate the area under the receiver-operating characteristic curve (AUC). We also determined the discrimination of metabolite measures discovered in the China Kadoorie Biobank and furthermore, their discrimination power when combined together with the metabolites discovered in our study and Framingham Stroke Risk Score. The analyses were performed in R version 3.2.5 (http://www.R-project.org/).

RESULTS

The baseline descriptive characteristics of study participants are shown in **Table 1**. In total there were 1,791 incident stroke events observed among 38,797 participants across the seven cohorts. The mean follow-up time ranged from 2 years in PROSPER, 6 years in the Rotterdam Study, and 7 years in EGCUT and FINRISK07 to 13 years in Whitehall II and 15 years in FINRISK97 and FHS.

The results of association analysis between circulating metabolites and incident stroke are shown in **Table 2**. The analysis revealed 27 significant metabolite associations in model 1. After further adjustment for hypertension status, diabetes and smoking, 7 metabolite associations survived correction for multiple testing. These included the amino

Table 1. Descriptive statistics of study population.

	Rotterdam Study		Whitehall II**		Finrisk97	
Variable*	Incident cases	Controls	Incident cases	Controls	Incident cases	Controls
N	257	2308	197	5792	474	6384
Age (years)	76.9 (6.2)	75.0 (6.1)	59.4 (5.9)	55.6 (6)	59.6 (10.4)	47.0 (12.9)
Women	54.1%	58%	25.4%	29.1%	38.6%	52.6%
Current Smoking	15.2%	13%	15.70%	9.40%	24.5%	23.7%
Diabetes	17.9%	14.3%	8.6%	4.4%	16.9%	4.9%
Hypertension	85.6%	81.0%	41.1%	28.0%	48.9%	21%
Systolic blood pressure (mmHg)	156.8 (23.9)	151.4 (20.1)	127.4 (16.3)	122.9 (16.5)	147.7 (22.3)	134.7 (19.2)
Diastolic blood pressure (mmHg)	79.7 (12.4)	79.2 (11.1)	78.3 (10.2)	77.5 (10.5)	86.1 (11.9)	81.9 (11.2)
Antihypertensive medication	51%	47.10%	21.8%	12.3%	27.4%	11.5%
BMI (kg/m2)	27.2 (3.5)	27.4 (4.2)	26.2 (4.2)	26 (3.9)	28.4 (4.8)	26.5 (4.5)
Follow-up time (years)	5.7 (3.5)	9.8 (3.5)	12.5 (4.9)	18.2 (3)	15.03 (4.2)	16.9 (3)
Total cholesterol	5.5 (0.98)	5.6 (0.98)	5.8 (1.1)	5.9 (1.1)	5.79 (1.1)	5.52 (1.1)
HDL cholesterol	1.4 (0.4)	1.5 (0.4)	1.5 (0.4)	1.5 (0.4)	1.29 (0.3)	1.41 (0.4)
LDL cholesterol	NA	NA	3.8 (1.01)	3.9 (0.9)	3.71 (0.9)	3.46 (0.9)
Triglycerides	NA	NA	1.3 (0.8)	1.4 (0.9)	1.79 (1.1)	1.46 (1.0)
Lipid lowering medication	21.4%	20.6%	5.1%	3.0%	7.8%	3.1%
Coronary Heart Disease	13.6%	10.8%	11.7%	5.9%	9.1%	1.9%
Stroke						
Hemorrhagic	32 (12.5%)	-	48 (24.4%)	-	69 (14.6%)	-
Ischemic	183 (71.2%)	-	126 (64.0%)	-	405 (85.4%)	-
Not defined	42 (16.3%)	-	23 (11.6%)	-	-	-

* Values are means ± standard deviation for continuous variables and percentages for dichotomous variables.
** While all other cohorts included participants of European ancestry, 87.3% of Whitehall II study cases were of European ancestry, 6.6% of Asian, 5.1% of African American and 1% of other.
***Lipid levels are expressed in mmol/l for all cohorts except for FHS (mg/dl).

acid histidine (hazard ratio (HR) per SD = 0.9, 95 % confidence interval (CI): 0.85, 0.94) and cholesterol in high-density lipoprotein (HDL) 2 (HR per SD = 0.91, 95% CI: 0.87, 0.97) which were associated with a decreased risk of stroke, glycolysis-related metabolite pyruvate (HR per SD = 1.1, 95% CI: 1.04, 1.14) and acute phase reaction markers glycoprotein acetyls (HR per SD = 1.09, 95% CI: 1.03, 1.15) which were associated with an increased risk of stroke, and several lipoprotein particles including HDL and low-density lipoprotein (LDL subfractions) (**Table 2, Figure 1**). Cholesterol in medium HDL was associated with decreased risk (HR per SD = 0.92, 95% CI: 0.87, 0.97), whereas triglycerides in medium and large LDL were associated with an increased risk of stroke (HR per SD = 1.09, 95% CI: 1.03, 1.14 and HR per SD = 1.09, 95% CI: 1.03, 1.14, respectively) (**Table 2, Figure 1**). The direction of effect across the cohorts showed no evidence of single cohort driving the associations (**Figure 2**). Whereas the Whitehall II study showed opposite

Finrisk07		PROSPER		EGCUT		FHS***	
Incident cases	Controls	Incident cases	Controls	Incident cases	Controls	Incident cases	Controls
107	4424	197	4627	308	10268	251	3203
62.0 (10.4)	51.9 (13.5)	75.9 (3.7)	75.2 (3.3)	66.3 (12.5)	44.5 (17.1)	58.1 (8.98)	51.7 (10.1)
42.1%	53.7%	54%	52.2%	54.9%	63.3%	47.4%	51.4%
21.5%	17.4%	28.90%	27.10%	17.5%	29.9%	24.9%	24.6%
15.9%	8.9%	18.3%	10.6%	35.4%	7.7%	15.9%	5.0%
43%	16.5%	58.9%	62.5%	66.2%	24.4%	60.2%	34.2%
149.8 (23.7)	136.4 (20.2)	157.1 (21.9)	154.5 (21.8)	142.8 (18.8)	125.7 (16.9)	137.8 (20.8)	126.2 (18.5)
83.1 (13.6)	79.3 (11.0)	84.6 (11.8)	83.7 (11.4)	83.4 (10.9)	77.6 (10.7)	81.6 (10.4)	78.9 (9.9)
34.6%	22.1%	70.6%	74.4%	69.5%	24.3%	36.3%	16.4%
28.0 (5)	27.16 (4.8)	26.5 (4.1)	26.9 (4.2)	29.1 (5.7)	26.4 (5.4)	27.6 (5.1)	26.7 (4.8)
7.25 (1.5)	7.75 (0.7)	1.9 (1.0)	3.3 (0.5)	6.9 (3.1)	8.9 (1.8)	14.7 (7)	22.4 (6.0)
5.23 (0.96)	5.28 (1)	5.64 (0.85)	5.68 (0.90)	6.0 (1.2)	5.7 (1.2)	215.4 (42.0)	205.1 (38.5)
1.44 (0.4)	1.44 (0.4)	1.25 (0.32)	1.28 (0.35)	1.5 (0.4)	1.6 (0.5)	47.4 (15.5)	49.7 (14.9)
3.11 (0.8)	3.20 (0.9)	3.77 (0.76)	3.79 (0.80)	2.5 (0.7)	2.3 (0.6)	138.2 (36.8)	131.4 (35.1)
1.49 (0.8)	1.42 (0.9)	1.57 (0.70)	1.54 (0.69)	1.9 (1.0)	1.6 (0.9)	150.9 (104.7)	123.5 (101.8)
25.2%	14.7%	52.3%	49.5%	13.6%	4.7%	6.0%	3.7%
5.6%	2.9%	16.8%	13.1%	35.1%	9.2%	12.4%	5.7%
23 (21.5%)	-	-	-	45 (14.6%)	-	30 (12%)	-
84 (78.5%)	-	-	-	261 (84.7%)	-	219 (87.3%)	-
-	-	-	-	11 (3.6%)	-	2 (0.8%)	-

direction of effect for apolipoprotein A, HDL, and HDL2 cholesterol, the findings showed a general spread for most HDL subfractions.

Table 2. Results of association analysis between incident stroke and metabolites.

			Model 1					Model 2		
Metabolite	**N**	**Ncases**	**HR**	**CI**	**P**	**N**	**Ncases**	**HR**	**CI**	**P**
Phenylalanine	35091	1527	1.11	1.06;1.17	4.88E-05	35036	1524	1.08	1.03;1.14	3.36E-03
Histidine*	35017	1526	0.89	0.84;0.93	7.94E-06	34962	1523	0.9	0.85;0.94	4.45E-05
plasma-ApoA1	35107	1529	0.91	0.86;0.96	7.14E-04	35052	1526	0.94	0.88;0.99	1.79E-02
HDL-cholesterol	35107	1529	0.89	0.84;0.94	2.89E-05	35052	1526	0.92	0.87;0.97	3.20E-03
HDL2-cholesterol*	35107	1529	0.88	0.84;0.93	9.13E-06	35052	1526	0.91	0.87;0.97	1.41E-03
IDL-triglycerides	38561	1780	1.1	1.05;1.16	6.06E-05	38494	1775	1.07	1.02;1.12	9.91E-03
LDL-triglycerides	35107	1529	1.12	1.06;1.18	3.93E-05	35052	1526	1.08	1.03;1.14	2.47E-03
Glucose	34980	1524	1.15	1.1;1.2	7.81E-11	34925	1521	1.06	1.01;1.11	1.87E-02
Lactate	35100	1529	1.12	1.07;1.18	1.11E-05	35045	1526	1.08	1.02;1.13	5.09E-03
Pyruvate*	24423	1205	1.13	1.08;1.18	1.37E-07	24368	1202	1.09	1.04;1.14	7.45E-04
Glycoprotein acetyls*	35101	1529	1.15	1.09;1.21	1.25E-07	35046	1526	1.09	1.03;1.15	1.27E-03
HDL-diametar	35107	1529	0.89	0.84;0.94	3.05E-05	35052	1526	0.92	0.87;0.98	6.73E-03
S-HDL-triglycerides	35108	1529	1.11	1.06;1.17	6.80E-05	35053	1526	1.07	1.01;1.12	1.97E-02
M-HDL-cholesterol*	38560	1780	0.89	0.85;0.94	2.07E-05	38493	1775	0.92	0.87;0.97	1.35E-03
M-HDL-cholesterol esters	35106	1529	0.9	0.85;0.95	2.05E-04	35051	1526	0.92	0.87;0.97	3.73E-03
M-HDL-free cholesterol	35106	1529	0.91	0.86;0.96	7.33E-04	35051	1526	0.93	0.88;0.98	8.24E-03
L-HDL-cholesterol	38555	1780	0.89	0.84;0.94	2.13E-05	38488	1775	0.92	0.88;0.98	5.50E-03
L-HDL-cholesterol esters	35101	1529	0.9	0.84;0.95	2.03E-04	35046	1526	0.93	0.88;0.99	1.37E-02
L-HDL-free cholesterol	35101	1529	0.89	0.84;0.94	1.25E-04	35046	1526	0.92	0.87;0.98	9.96E-03
L-HDL-total lipids	35101	1529	0.9	0.85;0.95	2.12E-04	35046	1526	0.93	0.88;0.99	1.70E-02
L-HDL-phospholipids	35101	1529	0.9	0.85;0.96	6.29E-04	35046	1526	0.94	0.89;1	3.49E-02
L-HDL concentration	35101	1529	0.9	0.85;0.96	8.53E-04	35046	1526	0.94	0.89;1	4.21E-02
XL-HDL-free cholesterol	35099	1527	0.91	0.86;0.96	8.31E-04	35044	1524	0.94	0.89;1	3.55E-02
S-LDL-triglycerides	29120	1332	1.12	1.06;1.18	2.95E-05	29065	1329	1.09	1.03;1.15	2.81E-03
L-LDL-triglycerides*	35107	1529	1.12	1.06;1.17	3.00E-05	35052	1526	1.09	1.03;1.14	1.67E-03
M-LDL-triglycerides*	35106	1529	1.12	1.06;1.18	1.68E-05	35051	1526	1.09	1.03;1.14	1.19E-03
XL-VLDL-triglycerides	38284	1769	1.09	1.04;1.14	1.56E-04	38217	1764	1.05	1;1.1	4.66E-02

Abbreviations: N - Total samples size; Ncases - Number of cases; HR - Hazard Ratio; 95% CI - 95% confidence interval; P - *p*-value; Model 1 - adjustment for age, gender, BMI, lipid-lowering medication and study-specific covariates if needed; Model 2- additional adjustment for smoking status, diabetes, and hypertension; *Associations that surpassed significance threshold in model 2.

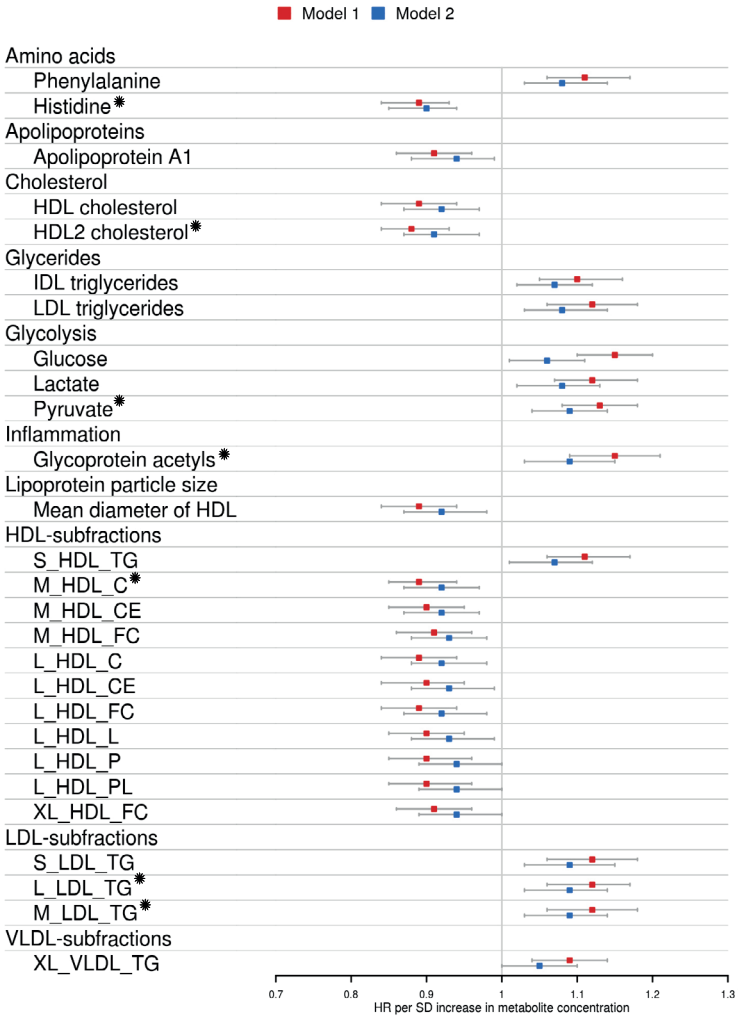

Figure 1. The results of association analysis between incident stroke and metabolites across two different models. Only associations that surpassed significance threshold are illustrated. The results are shown in red for model 1 and in blue for model 2. Hazard Ratios (HR) are denoted with boxes, while corresponding 95% confidence intervals of effect estimates are represented with whiskers. The associations that remained significant in model 2 are denoted by *.

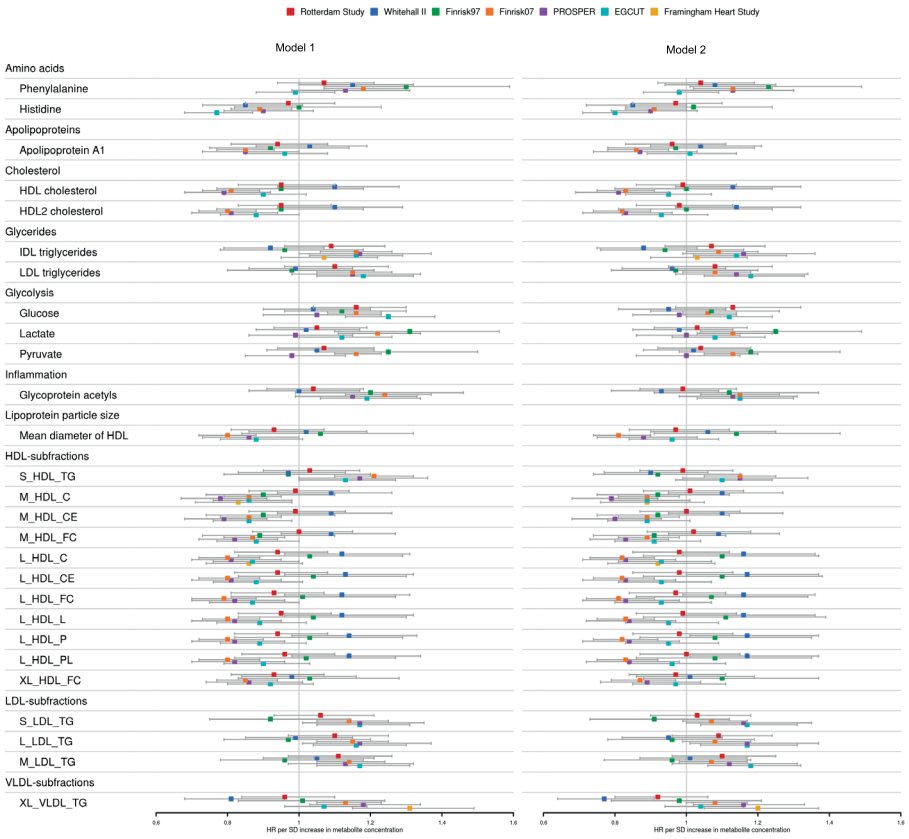

Figure 2. Associations that surpassed the threshold for multiple testing across two different models per study. The results for Rotterdam Study are shown in red, for Whitehall II in blue, for FINRISK97 in green, for FINRISK07 in orange, for PROSPER in purple, for EGCUT in turquoise and for FHS in golden. Hazard Ratios (HR) are denoted with boxes, while corresponding 95% confidence intervals of effect estimates are represented with whiskers.

When we stratified the analysis by stroke type, we observed differences between ischemic and hemorrhagic stroke events (**Table 3**). Amino acid histidine and cholesterol in HDL2 were associated with decreased risk of ischemic but not hemorrhagic incident stroke (**Table 3**). The differences were also observed for glycolysis-related metabolite pyruvate and acute phase reaction markers glycoprotein acetyls which were associated with increased risk of ischemic but not hemorrhagic stroke (**Table 3)**. Association between incident stroke events and LDL and HDL particles of various sizes was observed only in the overall analysis, suggesting contribution of both stroke subtypes (**Table 3**).

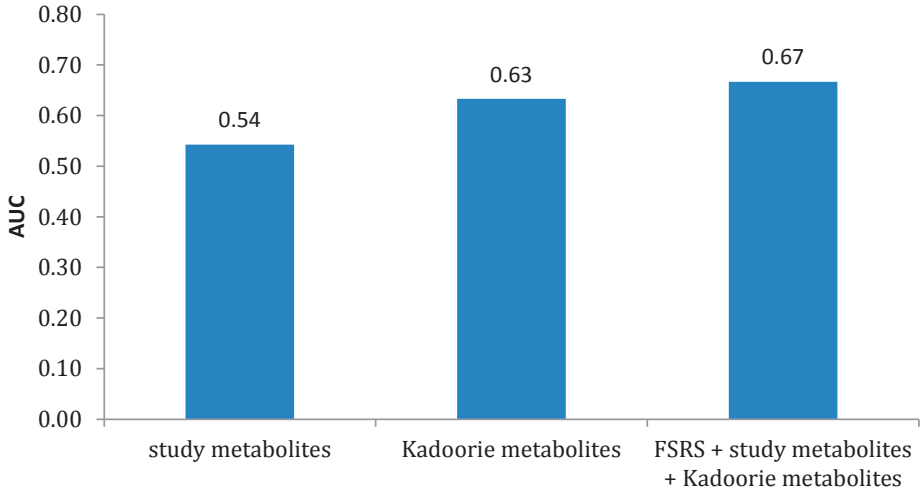

Figure 3. The area under the receiver-operating characteristic curve (AUC) in different prediction models. Study metabolites refer to circulating metabolites discovered in our study, while Kadoorie metabolites refer to metabolites discovered in the China Kadoorie Biobank. FSRS stands for Framingham Stroke Risk Score.

Furthermore, a significant association was observed between phenylalanine levels and increased risk of incident ischemic stroke (HR per SD = 1.12, 95% CI: 1.05, 1.19) and decreased risk of ischemic stroke and level of cholesterol in large HDL (HR per SD = 0.89, 95% CI: 0.84, 0.95) and of free cholesterol in the same particles (HR per SD = 0.89, 95% CI: 0.82, 0.95). There was no metabolite that surpassed the significant threshold in the analysis for hemorrhagic stroke.

The metabolites discovered in our study discriminate future stroke with the AUC of 0.54 (95% CI: 0.50, 0.58) (**Figure 3**). When analyses were repeated using the metabolites discovered in the China Kadoorie Biobank, the AUC was 0.63 (95% CI: 0.60, 0.67). Combining the metabolites discovered in our study and China Kadoorie Biobank, which is using the same metabolomics platform in Chinese population, together with Framingham Stroke Risk Score, lead to further improvement of the discrimination of future patients (AUC: 0.67, 95% CI: 0.63, 0.70) (**Figure 3**).

Table 3. Significant associations for overall incident stroke events and when classified by stroke type.

		Model 1					Model 2				
Metabolite	**Type**	N	Ncases	HR	CI	**P**	**N**	**Ncases**	**HR**	**CI**	**P**
Phenylalanine	all	35091	1527	1.11	1.06;1.17	**4.88E-05**	35036	1524	1.08	1.03;1.14	3.36E-03
	hemorrhagic	30144	214	0.94	0.81;1.09	3.90E-01	30092	214	0.91	0.78;1.05	2.00E-01
	ischemic*	30290	1051	1.16	1.09;1.23	**3.15E-06**	30236	1049	1.12	1.05;1.19	**4.13E-04**
Histidine	all*	35017	1526	0.89	0.84;0.93	**7.94E-06**	34962	1523	0.9	0.85;0.94	**4.45E-05**
	hemorrhagic	30070	214	0.95	0.82;1.1	4.78E-01	30018	214	0.96	0.83;1.11	6.16E-01
	ischemic*	30216	1050	0.88	0.82;0.94	**1.18E-04**	30162	1048	0.89	0.84;0.95	**4.94E-04**
Apolipoprotein A1	all	35107	1529	0.91	0.86;0.96	**7.14E-04**	35052	1526	0.94	0.88;0.99	1.79E-02
	hemorrhagic	30155	216	1.03	0.89;1.19	7.32E-01	30103	216	1.04	0.9;1.2	6.17E-01
	ischemic	30301	1051	0.89	0.83;0.95	**5.87E-04**	30247	1049	0.92	0.86;0.98	1.43E-02
HDL-cholesterol	all	35107	1529	0.89	0.84;0.94	**2.89E-05**	35052	1526	0.92	0.87;0.97	3.20E-03
	hemorrhagic	30155	216	1.04	0.9;1.21	5.63E-01	30103	216	1.07	0.92;1.24	3.97E-01
	ischemic	30301	1051	0.86	0.81;0.92	**1.82E-05**	30247	1049	0.9	0.84;0.96	1.89E-03
HDL2-cholesterol	all*	35107	1529	0.88	0.84;0.93	**9.13E-06**	35052	1526	0.91	0.87;0.97	**1.41E-03**
	hemorrhagic	30155	216	1.04	0.9;1.21	5.75E-01	30103	216	1.07	0.92;1.24	3.90E-01
	ischemic*	30301	1051	0.85	0.8;0.91	**2.85E-06**	30247	1049	0.89	0.83;0.95	**5.29E-04**
IDL-triglycerides	all	38561	1780	1.1	1.05;1.16	**6.06E-05**	38494	1775	1.07	1.02;1.12	9.91E-03
	hemorrhagic	33609	246	0.92	0.81;1.06	2.63E-01	33545	246	0.89	0.78;1.02	1.02E-01
	ischemic	33755	1270	1.13	1.07;1.2	**6.91E-06**	33689	1266	1.09	1.03;1.15	2.01E-03
LDL-triglycerides	all	35107	1529	1.12	1.06;1.18	**3.93E-05**	35052	1526	1.08	1.03;1.14	2.47E-03
	hemorrhagic	30155	216	1.02	0.88;1.18	8.27E-01	30103	216	0.99	0.85;1.14	8.62E-01
	ischemic	30301	1051	1.14	1.07;1.21	**2.89E-05**	30247	1049	1.1	1.04;1.17	1.82E-03
Glucose	all	34980	1524	1.15	1.1;1.2	**7.81E-11**	34925	1521	1.06	1.01;1.11	1.87E-02
	hemorrhagic	30033	214	1.13	0.99;1.28	7.07E-02	29981	214	1.09	0.96;1.24	2.20E-01

Table 3. Significant associations for overall incident stroke events and when classified by stroke type. (*continued*)

		Model 1					Model 2				
Metabolite	**Type**	N	Ncases	HR	CI	P	N	Ncases	HR	CI	P
	ischemic	30179	1048	1.17	1.12;1.23	**5.37E-11**	30125	1046	1.07	1.01;1.13	2.13E-02
Lactate	all	35100	1529	1.12	1.07;1.18	**1.11E-05**	35045	1526	1.08	1.02;1.13	5.09E-03
	hemorrhagic	30153	216	1.06	0.92;1.22	4.04E-01	30101	216	1.04	0.9;1.19	6.13E-01
	ischemic	30299	1051	1.16	1.09;1.24	**1.11E-06**	30245	1049	1.1	1.04;1.17	1.98E-03
Pyruvate	all*	24423	1205	1.13	1.08;1.18	**1.37E-07**	24368	1202	1.09	1.04;1.14	**7.45E-04**
	hemorrhagic	19481	167	0.99	0.84;1.16	8.70E-01	19429	167	0.96	0.81;1.12	5.94E-01
	ischemic*	19627	778	1.17	1.11;1.23	**2.86E-10**	19573	776	1.13	1.07;1.19	**1.93E-05**
Glycoprotein acetyls	all*	35101	1529	1.15	1.09;1.21	**1.25E-07**	35046	1526	1.09	1.03;1.15	**1.27E-03**
	hemorrhagic	30154	216	1.02	0.88;1.18	8.06E-01	30102	216	0.96	0.83;1.11	6.28E-01
	ischemic*	30300	1051	1.2	1.13;1.28	**8.55E-09**	30246	1049	1.13	1.06;1.2	**2.17E-04**
Mean diameter of HDL	all	35107	1529	0.89	0.84;0.94	**3.05E-05**	35052	1526	0.92	0.87;0.98	6.73E-03
	hemorrhagic	30155	216	1.03	0.89;1.2	6.98E-01	30103	216	1.07	0.92;1.24	4.08E-01
	ischemic	30301	1051	0.86	0.8;0.92	**1.28E-05**	30247	1049	0.9	0.84;0.96	2.93E-03
S-HDL-triglycerides	all	35108	1529	1.11	1.06;1.17	**6.80E-05**	35053	1526	1.07	1.01;1.12	1.97E-02
	hemorrhagic	30156	216	1	0.86;1.15	9.71E-01	30104	216	0.96	0.83;1.11	5.84E-01
	ischemic	30302	1051	1.14	1.08;1.22	**1.99E-05**	30248	1049	1.09	1.02;1.16	8.32E-03
M-HDL-cholesterol	all*	38560	1780	0.89	0.85;0.94	**2.07E-05**	38493	1775	0.92	0.87;0.97	**1.35E-03**
	hemorrhagic	33608	246	1	0.87;1.15	9.88E-01	33544	246	1.02	0.88;1.17	8.27E-01
	ischemic	33754	1270	0.88	0.83;0.93	**3.11E-05**	33688	1266	0.91	0.85;0.97	1.95E-03
M-HDL-cholesterol esters	all	35106	1529	0.9	0.85;0.95	**2.05E-04**	35051	1526	0.92	0.87;0.97	3.73E-03
	hemorrhagic	30154	216	0.99	0.86;1.14	8.84E-01	30102	216	1	0.87;1.16	9.72E-01
	ischemic	30300	1051	0.89	0.83;0.95	**4.60E-04**	30246	1049	0.91	0.86;0.98	6.99E-03
M-HDL-free cholesterol	all	35106	1529	0.91	0.86;0.96	**7.33E-04**	35051	1526	0.93	0.88;0.98	8.24E-03

Table 3. Significant associations for overall incident stroke events and when classified by stroke type. (*continued*)

		Model 1					Model 2				
Metabolite	**Type**	N	Ncases	HR	CI	P	N	Ncases	HR	CI	P
	hemorrhagic	30154	216	1.02	0.88;1.18	8.25E-01	30102	216	1.02	0.88;1.18	7.79E-01
	ischemic	30300	1051	0.9	0.84;0.96	1.75E-03	30246	1049	0.92	0.86;0.98	1.56E-02
L-HDL-cholesterol	all	38555	1780	0.89	0.84;0.94	**2.13E-05**	38488	1775	0.92	0.88;0.98	5.50E-03
	hemorrhagic	33603	246	1.07	0.93;1.24	3.19E-01	33539	246	1.11	0.96;1.28	1.56E-01
	ischemic*	33749	1270	0.85	0.8;0.91	**2.03E-06**	33683	1266	0.89	0.84;0.95	**9.00E-04**
L-HDL-cholesterol esters	all	35101	1529	0.9	0.84;0.95	**2.03E-04**	35046	1526	0.93	0.88;0.99	1.37E-02
	hemorrhagic	30149	216	1.07	0.92;1.24	3.88E-01	30097	216	1.1	0.95;1.28	2.13E-01
	ischemic	30295	1051	0.86	0.8;0.92	**4.28E-05**	30241	1049	0.9	0.84;0.97	3.59E-03
L-HDL-free cholesterol	all	35101	1529	0.89	0.84;0.94	**1.25E-04**	35046	1526	0.92	0.87;0.98	9.96E-03
	hemorrhagic	30149	216	1.09	0.93;1.26	2.81E-01	30097	216	1.12	0.96;1.3	1.44E-01
	ischemic*	30295	1051	0.85	0.79;0.91	**1.04E-05**	30241	1049	0.89	0.82;0.95	**1.33E-03**
L-HDL-total lipids	all	35101	1529	0.9	0.85;0.95	**2.12E-04**	35046	1526	0.93	0.88;0.99	1.70E-02
	hemorrhagic	30149	216	1.06	0.91;1.23	4.71E-01	30097	216	1.09	0.93;1.27	2.77E-01
	ischemic	30295	1051	0.87	0.81;0.93	**5.91E-05**	30241	1049	0.91	0.84;0.97	6.00E-03
L-HDL-concentration	all	35101	1529	0.9	0.85;0.96	**8.53E-04**	35046	1526	0.94	0.89;1	4.21E-02
	hemorrhagic	30149	216	1.09	0.94;1.27	2.73E-01	30097	216	1.12	0.96;1.3	1.44E-01
	ischemic	30295	1051	0.87	0.81;0.93	**1.30E-04**	30241	1049	0.91	0.85;0.98	1.04E-02
L-HDL-phospholipids	all	35101	1529	0.9	0.85;0.96	**6.29E-04**	35046	1526	0.94	0.89;1	3.49E-02
	hemorrhagic	30149	216	1.08	0.93;1.26	3.21E-01	30097	216	1.11	0.95;1.29	1.85E-01
	ischemic	30295	1051	0.87	0.81;0.93	**9.62E-05**	30241	1049	0.91	0.85;0.98	9.32E-03
XL-HDL-free cholesterol	all	35099	1527	0.91	0.86;0.96	**8.31E-04**	35044	1524	0.94	0.89;1	3.55E-02
	hemorrhagic	30147	216	1.07	0.93;1.24	3.52E-01	30095	216	1.09	0.94;1.26	2.38E-01

Table 3. Significant associations for overall incident stroke events and when classified by stroke type. (*continued*)

		Model 1					Model 2				
Metabolite	**Type**	N	Ncases	HR	CI	P	N	Ncases	HR	CI	P
	ischemic	30293	1049	0.88	0.82;0.94	**3.46E-04**	30239	1047	0.92	0.86;0.99	1.75E-02
S-LDL-triglycerides	all	29120	1332	1.12	1.06;1.18	**2.95E-05**	29065	1329	1.09	1.03;1.15	2.81E-03
	hemorrhagic	24168	168	1.04	0.88;1.22	6.44E-01	24116	168	1.01	0.86;1.19	8.73E-01
	ischemic	24314	925	1.14	1.07;1.21	**7.12E-05**	24260	923	1.09	1.03;1.17	5.79E-03
L-LDL-triglycerides	all*	35107	1529	1.12	1.06;1.17	**3.00E-05**	35052	1526	1.09	1.03;1.14	**1.67E-03**
	hemorrhagic	30155	216	1.01	0.87;1.17	9.12E-01	30103	216	0.98	0.85;1.13	7.87E-01
	ischemic	30301	1051	1.13	1.07;1.2	**4.39E-05**	30247	1049	1.1	1.03;1.17	2.20E-03
M-LDL-triglycerides	all*	35106	1529	1.12	1.06;1.18	**1.68E-05**	35051	1526	1.09	1.03;1.14	**1.19E-03**
	hemorrhagic	30154	216	1.04	0.9;1.2	5.70E-01	30102	216	1.02	0.88;1.17	8.35E-01
	ischemic	30300	1051	1.14	1.07;1.21	**2.84E-05**	30246	1049	1.1	1.04;1.17	1.80E-03
XL-VLDL-triglycerides	all	38284	1769	1.09	1.04;1.14	**1.56E-04**	38217	1764	1.05	1;1.1	4.66E-02
	hemorrhagic	33352	242	0.98	0.85;1.12	7.66E-01	33288	242	0.96	0.83;1.1	5.23E-01
	ischemic	33499	1263	1.12	1.06;1.18	**2.00E-05**	33433	1259	1.07	1.01;1.13	1.41E-02

Abbreviations: N - Total samples size; Ncases - Number of cases; HR - Hazard Ratio; 95% CI - 95% confidence interval; P - *p*-value; Model 1 - adjustment for age, gender, BMI, lipid-lowering medication and study-specific covariates if needed; Model 2- additional adjustment for smoking status, diabetes, and hypertension;
*Associations that surpassed significance threshold in model 2.

DISCUSSION

In this study, we identified ten metabolites associated with the risk of stroke. These include amino acid histidine, glycolysis-related metabolite pyruvate, acute phase reaction markers glycoprotein acetyls, cholesterol in HDL2, and lipoprotein subfractions such as cholesterol in medium HDL and triglycerides in medium and large LDL particles which showed association with incident stroke events. Amino acid phenylalanine and HDL subfractions including cholesterol and free cholesterol in large HDL were associated with ischemic incident stroke. This metabolite profile was independent of traditional risk factors including hypertension, diabetes, smoking, and BMI.

The strongest association was observed between amino acid histidine and risk of stroke. One SD increase in concentration of histidine was associated with 10% lower risk of stroke. The effect was very similar across studies, with only the Finrisk97 study showing no effect. Even though the same direction of effect was observed in both ischemic and hemorrhagic stroke subtype, the association was mainly driven by ischemic stroke. Histidine is a semi-essential amino acid as adults generally produce it while children may not. Histidine can be converted to histamine which shows a strong effect on vasodilatation and functions as a neurotransmitter in the brain.[28,29] Previous studies reported that oral administration of histidine can reduce blood pressure.[30-32] Plasma concentrations of histidine have been inversely associated with inflammation and oxidative stress in patients with chronic kidney disease and obese women with metabolic syndrome.[33-35] Furthermore, histidine has also been studied in relation to cerebral ischemia. Recent animal studies reported that histidine treatment remarkably alleviated the infarction induced by middle cerebral artery occlusion[36] and showed long term-neuroprotection after cerebral ischemia with decreased infarct volume and improved neurological function.[37] Even though our findings support the results of previous studies, in the most comprehensive study of stroke to date, histidine was not associated with ischemic and hemorrhagic stroke in individuals within the China Kadoorie Biobank. However, in the China Kadoorie Biobank, nominal association was found with myocardial infarction.[12] This could be explained either by environmental and ethnic differences of studied populations or difference in confounder adjusted for, in the present study we adjusted for more potential confounders including BMI, lipid-lowering medication, diabetes, and hypertension.

We also found the glycolysis-related metabolite, pyruvate, to be associated with increased risk of stroke. The analyses of stroke subtypes suggested that this association was driven by ischemic incident stroke events. Our findings suggested that 1 SD increase in pyruvate concentration was associated with 12% higher risk of ischemic stroke. Pyruvate is

the end-product of glycolysis and it is critical for supplying energy to the cell.[38] Pyruvate has previously been shown to protect against experimental stroke possibly by blocking inflammation.[39,40] In this light, our finding seems to contrast previously described effects of pyruvate. However, in a combined study of myocardial infarction and stroke using the same metabolomics platform as the present study, higher levels of pyruvate were also associated with a higher risk of cardiovascular disease.[41] The mechanism through which circulating level of pyruvate relates to stroke and cardiovascular disease is still to be elucidated.

Furthermore, acute phase marker glycoprotein acetyls mainly alpha-1 glycoprotein was associated with higher risk of stroke. The analyses revealed that the association was strongest for ischemic subtype, for which we found that an increase of 1 SD in the circulating compound was associated with 13% higher risk of ischemic stroke. Our results confirmed association of glycoprotein acetyl and ischemic stroke that was observed in individuals within the China Kadoorie Biobank.[12] Circulating levels of glycoprotein acetyls have previously been associated with cardiovascular diseases and dementia but also inflammatory disease, cancer, and mortality.[41-43]

Analyses focused on stroke subtypes revealed the association of essential amino acid phenylalanine with increased risk of ischemic stroke. One SD increase in concentration of phenylalanine was associated with 15% higher risk of ischemic stroke. Phenylalanine is a precursor for tyrosine and catecholamines including dopamine, epinephrine, and norepinephrine. Phenylalanine has previously been associated with risk of diabetes and cardiovascular disease.[41,44,45] As the association with phenylalanine remains after adjustment for diabetes, the association with stroke cannot be explained by impaired glucose tolerance. Phenylalanine did not associate with risk of hemorrhagic stroke.

Majority of circulating biomarkers measured by NMR metabolomics technology belong to lipid concentrations and composition of 14 lipoprotein subparticles. This provides an excellent opportunity for comprehensive investigation of lipoprotein particles in stroke, as the analyses of cholesterol and cholesterol subfraction has shown inconsistent results.[3,5,6] In our study population we observed association of cholesterol in medium HDL with decreased risk of stroke and triglycerides in large and medium LDL particles with increased risk of stroke. None of these lipoproteins measurements was found to be associated with stroke in the China Kadoorie Biobank.[12] Interestingly, the China Kadoorie Biobank reported association of low-, intermediate-, and low-density lipoproteins with ischemic stroke.[12] However, we were not able to confirm these results in our study population. Again, lack of replication could be explained by environmental and ethnic differences of studied populations or the confounders adjusted for.

Interestingly, using metabolites discovered in our study, we were able to discriminate future stroke with the AUC of 0.54. The metabolites discovered in the China Kadoorie Biobank also showed to be relevant for discriminating future stroke in our study population. Finally, when using metabolites discovered in our study and the China Kadoorie Biobank together with the Framingham Stroke Risk Score, we found further increase in the AUC. This suggests that the metabolites may have better utility for prediction of stroke and asks for use of other metabolomics platforms in order to discover additional metabolite measurements which could improve the risk prediction.

Strengths of our study are large sample size, prospective study design with detailed data collection over a long period of follow-up and the similar quantification method of circulating metabolites across the studies. Our study also has several limitations. With new improved methods available many other metabolites can be measured, which can be of importance to stroke.[46] Another limitation is differences in methods used across the cohorts to identify cases of incident stroke. As most of the cohorts used electronic health registries, this may have limited sensitivity which subsequently influenced power to identify novel significant associations. Furthermore, statistical power was also reduced in analyses of stroke subtypes as some of the cohorts were unable to distinguish between these. Another limitation is limited sample size for the analysis of hemorrhagic stroke which influenced our ability to detect novel associations for this stroke type.

To conclude, we found association of ten metabolites associated with risk of stroke in 1,791 incident stroke events observed among 38,797 individuals from seven population-based studies. The biological mechanisms underlying these associations should be subject of further studies.

REFERENCES

1. Mozaffarian. Heart Disease and Stroke Statistics-2015 Update: A Report From the American Heart Association (vol 131, pg e29, 2015). *Circulation* **131**, E535-E535 (2015).
2. Andersen, K.K., Olsen, T.S., Dehlendorff, C. & Kammersgaard, L.P. Hemorrhagic and ischemic strokes compared: stroke severity, mortality, and risk factors. *Stroke* **40**, 2068-72 (2009).
3. Goldstein, L.B. *et al.* Guidelines for the primary prevention of stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* **42**, 517-84 (2011).
4. Bos, M.J., Koudstaal, P.J., Hofman, A. & Ikram, M.A. Modifiable etiological factors and the burden of stroke from the Rotterdam study: a population-based cohort study. *PLoS Med* **11**, e1001634 (2014).
5. Zhang, Y. *et al.* Total and high-density lipoprotein cholesterol and stroke risk. *Stroke* **43**, 1768-74 (2012).
6. Amarenco, P., Labreuche, J. & Touboul, P.J. High-density lipoprotein-cholesterol and risk of stroke and carotid atherosclerosis: a systematic review. *Atherosclerosis* **196**, 489-96 (2008).
7. Saenger, A.K. & Christenson, R.H. Stroke biomarkers: progress and challenges for diagnosis, prognosis, differentiation, and treatment. *Clin Chem* **56**, 21-33 (2010).
8. Makin, S.D., Doubal, F.N., Dennis, M.S. & Wardlaw, J.M. Clinically Confirmed Stroke With Negative Diffusion-Weighted Imaging Magnetic Resonance Imaging: Longitudinal Study of Clinical Outcomes, Stroke Recurrence, and Systematic Review. *Stroke* **46**, 3142-8 (2015).
9. Jung, J.Y. *et al.* 1H-NMR-based metabolomics study of cerebral infarction. *Stroke* **42**, 1282-8 (2011).
10. Wang, D., Kong, J., Wu, J., Wang, X. & Lai, M. GC-MS-based metabolomics identifies an amino acid signature of acute ischemic stroke. *Neurosci Lett* **642**, 7-13 (2017).
11. Lee, Y., Khan, A., Hong, S., Jee, S.H. & Park, Y.H. A metabolomic study on high-risk stroke patients determines low levels of serum lysine metabolites: a retrospective cohort study. *Mol Biosyst* **13**, 1109-1120 (2017).
12. Holmes, M.V. *et al.* Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. *J Am Coll Cardiol* **71**, 620-632 (2018).
13. Wieberdink, R.G., Ikram, M.A., Hofman, A., Koudstaal, P.J. & Breteler, M.M.B. Trends in stroke incidence rates and stroke risk factors in Rotterdam, the Netherlands from 1990 to 2008. *European Journal of Epidemiology* **27**, 287-295 (2012).
14. Kivimaki, M. *et al.* Validity of Cardiovascular Disease Event Ascertainment Using Linkage to UK Hospital Records. *Epidemiology* **28**, 735-739 (2017).
15. Britton, A. *et al.* Validating self-reported strokes in a longitudinal UK cohort study (Whitehall II): Extracting information from hospital medical records versus the Hospital Episode Statistics database. *Bmc Medical Research Methodology* **12** (2012).
16. Shepherd, J. *et al.* The design of a Prospective Study of Pravastatin in the Elderly at Risk (PROSPER). *American Journal of Cardiology* **84**, 1192-1197 (1999).
17. Carandang, R. *et al.* Trends in incidence, lifetime risk, severity, and 30-day mortality of stroke over the past 50 years. *JAMA* **296**, 2939-46 (2006).
18. Seshadri, S. *et al.* The lifetime risk of stroke: estimates from the Framingham Study. *Stroke* **37**, 345-50 (2006).
19. Soininen, P., Kangas, A.J., Wurtz, P., Suna, T. & Ala-Korpela, M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics. *Circulation-Cardiovascular Genetics* **8**, 192-206 (2015).

20. Soininen, P. *et al.* High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* **134**, 1781-1785 (2009).
21. Freedman, D.S. *et al.* Sex and age differences in lipoprotein subclasses measured by nuclear magnetic resonance spectroscopy: the Framingham Study. *Clin Chem* **50**, 1189-200 (2004).
22. Kathiresan, S. *et al.* A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet* **8 Suppl 1**, S17 (2007).
23. Delles, C. *et al.* Nuclear magnetic resonance-based metabolomics identifies phenylalanine as a novel predictor of incident heart failure hospitalisation: results from PROSPER and FINRISK 1997. *European Journal of Heart Failure* **20**, 663-673 (2018).
24. Akbaraly, T. *et al.* Association of circulating metabolites with healthy diet and risk of cardiovascular disease: analysis of two cohort studies. *Scientific Reports* **8**(2018).
25. Kettunen, J. *et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* **7**, 11122 (2016).
26. van der Lee, S.J. *et al.* Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies. *Alzheimers Dement* (2018).
27. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221-227 (2005).
28. Ignesti, G. *et al.* Increased desensitization by picomolar phorbol ester of the endothelium-mediated effect of histamine in the perfused rat mesenteric bed. *Inflamm Res* **45**, 171-5 (1996).
29. Wang, L. *et al.* Plasma Amino Acid Profile in Patients with Aortic Dissection. *Sci Rep* **7**, 40146 (2017).
30. Toba, H. *et al.* Oral L-histidine exerts antihypertensive effects via central histamine H3 receptors and decreases nitric oxide content in the rostral ventrolateral medulla in spontaneously hypertensive rats. *Clin Exp Pharmacol Physiol* **37**, 62-8 (2010).
31. Tuttle, K.R., Milton, J.E., Packard, D.P., Shuler, L.A. & Short, R.A. Dietary amino acids and blood pressure: a cohort study of patients with cardiovascular disease. *Am J Kidney Dis* **59**, 803-9 (2012).
32. Jennings, A. *et al.* Amino Acid Intakes Are Inversely Associated with Arterial Stiffness and Central Blood Pressure in Women. *J Nutr* **145**, 2130-8 (2015).
33. Watanabe, M. *et al.* Consequences of low plasma histidine in chronic kidney disease patients: associations with inflammation, oxidative stress, and mortality. *Am J Clin Nutr* **87**, 1860-6 (2008).
34. Mihalik, S.J. *et al.* Metabolomic profiling of fatty acid and amino acid metabolism in youth with obesity and type 2 diabetes: evidence for enhanced mitochondrial oxidation. *Diabetes Care* **35**, 605-11 (2012).
35. Feng, R.N. *et al.* Histidine supplementation improves insulin resistance through suppressed inflammation in obese women with the metabolic syndrome: a randomised controlled trial. *Diabetologia* **56**, 985-994 (2013).
36. Adachi, N., Liu, K. & Arai, T. Prevention of brain infarction by postischemic administration of histidine in rats. *Brain Res* **1039**, 220-3 (2005).
37. Liao, R.J. *et al.* Histidine provides long-term neuroprotection after cerebral ischemia through promoting astrocyte migration. *Scientific Reports* **5**(2015).
38. Gray, L.R., Tompkins, S.C. & Taylor, E.B. Regulation of pyruvate metabolism and human disease. *Cell Mol Life Sci* **71**, 2577-604 (2014).
39. Wang, Q. *et al.* Pyruvate protects against experimental stroke via an anti-inflammatory mechanism. *Neurobiol Dis* **36**, 223-31 (2009).
40. Ryou, M.G. *et al.* Pyruvate protects the brain against ischemia-reperfusion injury by activating the erythropoietin signaling pathway. *Stroke* **43**, 1101-7 (2012).

41. Wurtz, P. *et al.* Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* **131**, 774-85 (2015).
42. Connelly, M.A., Gruppen, E.G., Otvos, J.D. & Dullaart, R.P.F. Inflammatory glycoproteins in cardio-metabolic disorders, autoimmune diseases and cancer. *Clin Chim Acta* **459**, 177-186 (2016).
43. Lawler, P.R. *et al.* Circulating N-Linked Glycoprotein Acetyls and Longitudinal Mortality Risk. *Circ Res* **118**, 1106-15 (2016).
44. Wang, T.J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat Med* **17**, 448-53 (2011).
45. Floegel, A. *et al.* Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. *Diabetes* **62**, 639-648 (2013).
46. Wishart, D.S. *et al.* HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* **41**, D801-7 (2013).