OPEN ACCESS

Metabolic phenotyping and cardiovascular disease: an overview of evidence from epidemiological settings

Aikaterini Iliou,¹ Emmanuel Mikros,¹ Ibrahim Karaman,² Freya Elliott,³ Julian L Griffin,⁴ Ioanna Tzoulaki,^{2,5,6} Paul Elliott ^{2,6,7,8}

¹Pharmacy, National and Kapodistrian University of Athens School of Health Sciences, Athens, Attica, Greece

²Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

³School of Medicine and Dentistry, Queen Mary University, London, UK

⁴Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK

⁵Department of Hygiene and Epidemiology, University of Ioannina, Ioannina, Greece

⁶BHF Research Centre for Excellence, Faculty of Medicine, Imperial College London, London, UK

⁷MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK

⁸Imperial College Biomedical Research Centre, Imperial College London, London, UK

Correspondence to

Professor Paul Elliott, MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College of Science Technology and Medicine, London W2 1PG, UK; p.elliott@imperial.ac.uk

Received 15 October 2020

Revised 5 January 2021

Accepted 8 January 2021

Published Online First

19 February 2021

© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY. Published by BMJ.

To cite: Iliou A, Mikros E, Karaman I, et al. *Heart* 2021;**107**:1123–1129.

ABSTRACT

Metabolomics, the comprehensive measurement of low-molecular-weight molecules in biological fluids used for metabolic phenotyping, has emerged as a promising tool to better understand pathways underlying cardiovascular disease (CVD) and to improve cardiovascular risk stratification. Here, we present the main methodologies for metabolic phenotyping, the methodological steps to analyse these data in epidemiological settings and the associated challenges. We discuss evidence from epidemiological studies linking metabolites to coronary heart disease and stroke. These studies indicate the systemic nature of CVD and identify associated metabolic pathways such as gut microbial cometabolism, branched-chain amino acids, glycerophospholipid and cholesterol metabolism, as well as activation of inflammatory processes. Integration of metabolomic with genomic data can provide new evidence for involved biochemical pathways and potential for causality using Mendelian randomisation. The clinical utility of metabolic biomarkers for cardiovascular risk stratification in healthy individuals has not yet been established. As sample sizes with high-dimensional molecular data increase in epidemiological settings, integration of metabolomic data across studies and platforms with other molecular data will lead to new understanding of the metabolic processes underlying CVD and contribute to identification of potentially novel preventive and pharmacological targets. Metabolic phenotyping offers a powerful tool in the characterisation of the molecular signatures of CVD, paving the way to new mechanistic understanding and therapies, as well as improving risk prediction of CVD patients. However, there are still challenges to face in order to contribute to clinically important improvements in CVD.

INTRODUCTION

Cardiovascular disease (CVD) remains the major cause of death globally, with an estimated 85 million deaths annually.¹ CVD comprises a diverse set of diseases, with coronary heart disease (CHD) and stroke being the most common manifestations. A large body of research over the past decades has identified a range of well-established risk factors, including male sex, high blood pressure, high total and low-density lipoprotein (LDL) cholesterol levels, smoking and type 2 diabetes, as well as many genetic factors that affect an individual's risk of developing CVD.² Nonetheless, the molecular mechanisms linking these and other risk factors to CVD are only partly understood, particularly how molecular mechanisms interact with

environmental exposures such as diet and xenobiotics. Better understanding of the underlying pathways which are implicated in CVD pathophysiology is important to design novel or improved strategies for prevention, risk stratification and treatment.

Metabolomics, the comprehensive measurement of low-molecular-weight molecules in biological fluids, provides an assessment of the metabolic signatures (metabolic phenotype) of intrinsic and extrinsic exposures (the internal and external exposures) from a variety of sources, including genetic, dietary, lifestyle, gut microbial and psychosocial factors. This involves the biochemical analyses of multiple metabolites in biological fluids, tissue homogenates as well as intact tissues. Substantial progress has been made in metabolomics through improved instrument performance and use of advanced chemoinformatic and bioinformatic tools for data acquisition, processing and analysis. These have facilitated the investigation of metabolites in relation to CVD in large population studies. Applications of metabolic phenotyping in CVD have so far been multifaceted, including the use of metabolic biomarkers to study the effects of lifestyle and environmental exposures on CVD risk,^{3,4} investigation of mechanisms and pathophysiological processes underlying CVD development,^{5,6} and progression and evaluation of prognostic biomarkers.^{7–9} Here, we provide a brief account of metabolomics technologies and their application in CVD research, focusing on epidemiological studies of CVD and its main components, CHD and stroke.

METABOLOMICS APPROACHES

There is no single technology to measure the entire metabolome. To address this, a wide variety of approaches, biological fluids, analytical platforms and statistical methods have been proposed to systematically increase the coverage of the metabolome. Analytical advances have allowed the simultaneous detection and quantitation of a variety of small molecules with different chemical properties and structures, including carbohydrates, lipids, steroids, organic acids, amino acids, peptides, energy-related metabolites and gut microbial metabolites, ranging from aqueously soluble metabolites to non-polar lipids and lipophilic components.

The application of metabolomics can be hypothesis driven, through targeted (closed) methods quantifying a small number of structurally related metabolites, or an untargeted (open, hypothesis generating) approach where the aim is to cover as

much of the metabolome as possible. With the targeted method, only a preselected group of metabolites that usually are chemically related, belonging to a metabolic class or pathway, is measured. With the use of standards labelled with stable isotopes, these assays can be made quantitative. In untargeted approaches, the chemical identity of the metabolites may not be known *a priori*, and chemical/spectral annotation may be needed post hoc to identify the molecular species. Furthermore, the untargeted methods are not optimised for a particular class of compounds and, thus, are often less sensitive than targeted methods and may be semiquantitative rather than fully quantitative.

Samples

Many known metabolites are present in blood and urine, which are the most frequently used sample types in human metabolic phenotyping. They are relatively easy to collect and store in epidemiological settings and carry extensive information on the metabolic phenotype. Blood is sampled as either blood plasma or serum, and contains a wide range of aqueously soluble metabolites and lipids, with the latter largely contained within lipoprotein fractions. Urine is a biological waste material containing products of endogenous metabolism, gut microbial cometabolism, renal function and exogenous exposures, and therefore also captures a complex and wide set of metabolic processes.¹⁰ It also has the advantage that it can significantly concentrate waste products of metabolic pathways. Other biofluids that have been used include saliva and cerebrospinal fluid, tissue extracts from atherosclerotic plaques or cell extracts. The latter may reveal tissue-specific biological processes that cannot be captured in blood and urine. Variations in sample collection procedures, handling, transport and storage both within and between studies may affect the metabolic profile and introduce bias, highlighting the need for standard operating procedures and extensive quality control (QC).^{11 12}

Analytical platforms and analysis workflow

The majority of metabolic phenotyping data are acquired using nuclear magnetic resonance (NMR) spectroscopy, and mass

spectrometry (MS)-hyphenated instrumental platforms, mainly liquid chromatography–MS and gas chromatography–MS. The relative advantages and disadvantages of each technology and their main applications are described in [table 1](#).

Metabolomics analysis can be divided into four main steps: preprocessing, statistical analysis, identification of the unknown features (for untargeted analyses) and pathway analysis ([figure 1](#)). Postdata acquisition processing (so-called preprocessing) in metabolomics is complex and requires a relatively sophisticated methodology as datasets can be very large as a result of the high-throughput nature of the data.¹³ QC samples, often pooled across samples included in the study, are used to assess reproducibility and to account for instrument drift, analytical signal variability, and baseline and retention time shifts across the same (intra-) or different (inter-) batches.¹⁴

Statistical analysis needs to take into account characteristics of metabolomics data, including a high degree of collinearity, high-dimensionality, non-linearity, missingness and non-normality.¹⁵ Both univariate and multivariate methods are used to address different types of metabolomics data and study questions. In univariate approaches, standard generalised linear models are commonly used. To account for multiple hypothesis testing, Bonferroni, false discovery rate correction and a less stringent metabolome-wide significance-level approach¹⁶ are often reported. Multivariate approaches are also used to tease out latent information from spectral data and to select relevant metabolic features. The most common multivariate approaches fall within two groups: unsupervised, such as principal component analysis, and supervised, such as partial least squares discriminant analysis. Unsupervised methods involve data reduction to differentiate clusters of samples that share common variation, while supervised methods focus on group separation in the data ([figure 2](#)).

Unknown statistically significant metabolic features need to be identified using analytical chemistry and bioinformatics analyses to relate features in the spectra to individual metabolites, often making use of online databases (eg, LipidMaps, Metlin and the Human Metabolome Database). The Metabolomics Society has

Table 1 Strengths and weaknesses of the core instrumentation used in metabolomics

	NMR	GC-MS	LC-MS/MS Q-TOF/Orbitrap	LC-MS/MS, triple quadrupole
Applications	Targeted and untargeted	Targeted and untargeted	Targeted and untargeted	Targeted
Strengths	<ul style="list-style-type: none"> ▶ Non-destructive technique. ▶ High reproducibility. ▶ Quantitative analysis. ▶ Relatively easy metabolite identification. 	<ul style="list-style-type: none"> ▶ Robust and reproducible metabolite separation. ▶ Available databases to aid metabolite identification, for example, NIST. ▶ Quantitative analysis. ▶ Less expensive compared with other MS platforms. 	<ul style="list-style-type: none"> ▶ High resolution. ▶ Accurate mass determination. ▶ Thousands of metabolites detected in a single run. ▶ Fragment fingerprinting using MSⁿ experiments. ▶ Derivatisation is rarely needed, though ionisation of the molecules is required. 	<ul style="list-style-type: none"> ▶ High specificity when combined with retention time. ▶ Quantitative analysis with the use of stable isotope labelled standards. ▶ Direct flow injection can be used with short run times, suited for targeted analysis of chemically related compounds. ▶ Derivatisation is rarely needed, though ionisation of the molecules is required.
Challenges	<ul style="list-style-type: none"> ▶ Low sensitivity compared with MS. ▶ Large sample volumes required compared with MS. ▶ Up to 200 metabolites can be detected in a single run due to overlapping signals compared with thousands detected in MS applications. 	<ul style="list-style-type: none"> ▶ Destructive technique. ▶ Prior derivatisation process is needed. ▶ Unsuitable for metabolites that thermally decompose. 	<ul style="list-style-type: none"> ▶ Destructive technique. ▶ Relatively expensive compared with other MS techniques. ▶ Complex metabolite identification. ▶ Noise and redundancy in untargeted data. ▶ Lack of absolute quantification in untargeted applications. ▶ Low reproducibility, requiring quality control samples. 	<ul style="list-style-type: none"> ▶ Destructive technique. ▶ Hypothesis driven. ▶ Difficult method development.

GC-MS, gas chromatography–mass spectrometry; LC-MS/MS, liquid chromatography–tandem mass spectrometry; MS, mass spectrometry; NIST, National Institute of Standards and Technology; NMR, nuclear magnetic resonance; Q-TOF, quadrupole time-of-flight mass spectrometer.

Figure 1 Schematic of the metabolic phenotyping analysis workflow. LC-MS, liquid chromatography–mass spectrometry; NMR, nuclear magnetic resonance; QC, quality control.

proposed a scoring system that describes the confidence of the metabolite identity assignment.¹⁷ The list of identified metabolites is often followed up by pathway analyses to aid interpretation of the results by placing the identified metabolites into relevant biochemical pathways (figure 3).^{18 19}

The nature of metabolomics data and the use of different analytical techniques and platforms for metabolite measurement and quantification pose special challenges to standardisation and harmonisation across studies and laboratories. Different preprocessing strategies and lack of information on metabolite identification and quantification procedures may introduce heterogeneity between studies. Several initiatives are attempting to synthesise metabolomics data from different studies such as the Consortium of Metabolomics Studies, which has focused mainly to date on pooling data from single platforms.²⁰ Extension of these efforts to combine data across different assay platforms is much more challenging but is expected to enhance the generalisability, reproducibility and molecular scope of metabolomic findings and to facilitate their implementation in clinical practice.

Figure 2 Schematic representation of score plots from (A) PCA and (B) partial least squares discriminant analysis models on a hypothetical two-group data where between-group variation is smaller than sample-to-sample variation. Within-group variation dominates; therefore, the group pattern is better revealed by separation using supervised methods. PC, principal component; LV, latent variable.

Applications of metabolic phenotyping in CVD

Epidemiological studies employing metabolic phenotyping in CVD have focused mainly on the identification of novel biomarkers of CVD, exploring biochemical pathways that underly cardiovascular phenotypes, Mendelian randomisation (MR) to explore causality and assessment of metabolomics in CVD risk prediction.

Early epidemiological studies on metabolomics and CVD outcomes were predominantly cross-sectional investigations reporting metabolites that were mainly consequences of clinical disease. With improved high-throughput technologies, data from prospective investigations have since been generated (table 2 summarises their study design and main findings). We discuss some of the major results from these studies further.

Lipid metabolic pathways

The evidence linking lipid metabolism and dyslipidaemia with CVD is well established. Metabolic profiling has further provided an in-depth characterisation of lipid species in MS technologies and has highlighted the potential role of specific lipid metabolic pathways such as sphingolipids, glycerophospholipids and glycerolipids as biomarkers of disease.

In a large-scale MS-based metabolomics study across three cohort studies, Ganna *et al*⁹ showed that after adjustment for traditional CVD risk factors, four lipid-related metabolites were associated with CHD: lysophosphatidylcholine (LPC) 18:1, LPC 18:2 and sphingomyelin 28:1 (inverse associations) and monoglyceride (MG) 18:2 (direct association). LPCs were also inversely associated with body mass index and presence of subclinical CVD cross-sectionally, while a reverse pattern was observed for MG 18:2. MG 18:2 is central in the synthesis and breakdown of triglycerides, while MGs are precursors for diglycerides, which are known to inhibit insulin signalling acting through protein kinase C ϵ , mechanistically linking excess lipid storage with insulin resistance.²¹ An inverse association between LPC 18:2 and 17:O in relation to incident myocardial infarction was also reported by another prospective multicohort investigation.⁸ LPCs are mainly contained in high-density lipoproteins (HDLs) and are produced on hydrolysis of phosphatidylcholines (PCs) in acute and chronic inflammation. Their association with CVD outcomes in these studies add further support to the importance of systemic inflammatory processes in the pathogenesis of CVD.

Figure 3 Ingenuity pathway analysis network showing molecular relationships between metabolites which have been associated with atherosclerosis (blue: inverse association, orange: direct association) (adapted from Tzoulaki *et al.*⁶). ;AARS, alanyl-tRNA synthetase; ALT, alanine transaminase; CPO, carboxypeptidase O; HTT, Huntington disease protein; LDL, low-density lipoprotein; LPO, lactoperoxidase; SFTPD, surfactant protein D.

Several studies have highlighted the importance of glycerophospholipids, mainly sphingomyelins, in CVD. Among them, sphingomyelin 32:1 has been found to be directly associated with incident ischaemic stroke, forearm vasodilation resistance (a marker of subclinical CVD) and inversely with leucocyte count (a marker of inflammation); however, these associations were attenuated when adjusting for traditional risk factors.²² Sphingomyelins are found in plasma and cell membranes and plasma lipoproteins that are involved in apoptosis and oxidative stress and have been associated with atherosclerotic plaque instability and lipid plaque burden.²³ Sphingomyelins can be broken down to ceramides, another important class of signalling molecules which may be involved in the pathogenesis of CVD through multiple pathways including lipotoxicity, inflammation and apoptosis.²⁴ Studies in animal models have suggested that pharmacological inhibition or genetic ablation of enzymes driving ceramide synthesis ameliorate atherosclerosis and related pathways.²⁵ However, results from epidemiological studies are inconsistent, with some but not all studies showing a direct association between different ceramides and future CVD.²⁶

Neutral lipids, in particular, cholesterol esters and triacylglycerols (TAGs) with a low carbon number and double-bond content have been associated with CVD in a prospective cohort study.²⁷ TAGs are associated with hepatic steatosis and raised de

novo lipogenesis, in part as a consequence of diets high in carbohydrates, in human feeding studies.²⁸

Epidemiological studies have also investigated the role of different lipid particles and lipoproteins through ¹H NMR measurements. For example, very low-density lipoprotein (VLDL), intermediate-density lipoprotein (IDL) and LDL, as well as triglycerides in all lipoproteins (mostly HDL particles) were directly associated with myocardial infarction and ischaemic stroke, while cholesterol in large HDL (but not in small) was inversely associated with these outcomes.²⁹ In another prospective investigation of 105 different NMR measured lipoproteins,⁵ lower-density lipoproteins showed strong direct associations with atherosclerosis and CVD with total apolipoprotein B and apolipoprotein B within total plasma LDL being associated with CVD after adjustment for non-lipid CVD risk factors.

Other metabolic pathways

Other prospective studies have identified an NMR-based metabolic fingerprint associated with CVD. In Würtz *et al.*,³⁰ four metabolites were associated with incident CVD: phenylalanine and monounsaturated fatty acid levels were directly associated with incident CVD, while omega-6 fatty acids and docosahexaenoic acid levels were associated with lower risk of CVD. Vaarhorst

Table 2 Summary of prospective epidemiological studies examining associations between metabolites and CVD

Reference	Study design	Cases (N)	Metabolomics platform	Main results
Ganna <i>et al</i> ⁸	Prospective multicohort: TwinGene, ULSAM and PIVUS	131 CHD events in discovery, 282 events in replication	LC-MS untargeted approach	LPC 18:1, LPC 18:2, sphingomyelin 28:1 and MG 18:2 associated with CHD MG 18:2 showed weak evidence for a causal effect on CHD risk using Mendelian randomisation analysis
Stegemann <i>et al</i> ²⁷	Prospective cohort study: Bruneck study	90 CVD events	LC-MS shotgun lipidomics, 135 lipid species	Individual species of cholesterol esters, LPCs, phosphatidylcholines, phosphatidylethanolamines, sphingomyelins and triacylglycerols were associated with CVD. Small improvement in discrimination over standard CVD risk factors for disease prediction
Würtz <i>et al</i> ³⁰	Prospective multicohort: FINRISK study, SABRE study and British Women's Health and Heart Study	800 CVD events in discovery, 573 and 368 events in replication	NMR and targeted LC-MS with Metabolon 68 NMR metabolites	Phenylalanine, monounsaturated fatty acids, omega-6 fatty acids, and docosahexaenoic acids were associated with CVD events Minimal improvement in discrimination over standard CVD risk factors for disease prediction
Ward-Caviness <i>et al</i> ⁶	Prospective multicohort: KORA S4, KORA S2 and AGES-REFINE	n=67, 112, 87 myocardial infarction cases, respectively	Targeted Biocrates AbsoluteIDQ p150 Kit LC-MS 188 metabolites	Arginine and LPC (LPC 17:0 and LPC 18:2) associated with myocardial infarction Metabolites increased the predictive value of the Framingham Risk Score in KORA S4 (from 0.70 to 0.78).
Holmes <i>et al</i> ²⁹	Nested case-control study: China Kadoorie Biobank	912 myocardial infarctions, 1146 ischaemic strokes, 1138 intracerebral haemorrhage	NMR 225 metabolites	Triglycerides within all lipoproteins showed similar associations with myocardial infarction and ischaemic stroke, but not with intracerebral haemorrhage. Glycoprotein acetyls, ketone bodies, glucose and docosahexaenoic acid were associated with all three diseases.
Lind <i>et al</i> ²²	Prospective multicohort: TwinGene, ULSAM and PIVUS	177 cases stroke events in discovery, 194 stroke events in replication	LC-MS untargeted approach	Sphingomyelin 32:1 was associated with incident stroke.
Tzoulaki <i>et al</i> ⁵	Cross-sectional and prospective multicohort: MESA, Rotterdam Study and LOLIPOP	630 CVD events	Untargeted NMR	25 metabolites of atherosclerosis also associated with CVD, associations attenuated after adjustments for risk factors

AGES-REFINE, Age, Gene/Environment Susceptibility (AGES) and the Risk Evaluation For Infarct Estimates (REFINE)-Reykjavik studies; CHD, coronary heart disease; CVD, cardiovascular disease; FINRISK, Finland Cardiovascular Risk Study; KORA, Collaborative Health Research in the Region of Augsburg; LC-MS, liquid chromatography-mass spectrometry; LOLIPOP, London Life Sciences Prospective Population Study; LPC, lysophosphatidylcholines; MESA, Multi-Ethnic Study of Atherosclerosis; MG, monoglyceride; NMR, nuclear magnetic resonance; PIVUS, Prospective Investigation of the Vasculature in Uppsala Seniors; SABRE, Southall And Brent Revisited Study; ULSAM, Uppsala Longitudinal Study of Adult Men.

*et al*³¹ used the least absolute shrinkage and selection operator (LASSO) algorithm to identify a metabolite score consisting of 13 ¹H NMR signals associated with CHD, comprising a lipid fraction, glucose, valine, ornithine, glutamine, creatine, glycoproteins, citrate and 1,5-anhydrosorbitol. Meanwhile, Tzoulaki *et al*⁵ evaluated the association of ¹H NMR features with coronary artery calcium and intima media thickness. They found a range of metabolites associated with atherosclerosis, including alanine, glycine, methionine, glucose, acetaminophen-glucuronide, glycerol, acetyl glycoproteins, myo-inositol, mannose, 1,5-anhydrosorbitol, glutamate, glutamine, N,N-dimethylglycine, lysine, phenylalanine, 5-oxoproline, 3-hydroxybutyrate, citrate and albumin. Several of these metabolites were also associated with incident CVD, highlighting the importance of these pathways in progression to clinical CVD.

The aforementioned analyses highlight the interactions of amino acids with CVD, in addition to the well-established links of carbohydrates and lipid metabolism with CVD. While further work is needed to understand the role of specific amino acids in CVD, there may be shared mechanisms with insulin resistance and type 2 diabetes, where studies have shown that branched chain amino acids, glycine and phenylalanine contribute to insulin resistance.^{32,33}

Overall, metabolomics studies to date have revealed systemic disturbances and interconnected pathways underlying CVD. These pathways include lipid, fatty acid and carbohydrate metabolism, branched chain amino acids and aromatic acid metabolism, tricarboxylic acid and urea cycle, and muscle metabolism (figure 3). The identified pathways point to the activation of inflammatory processes and oxidative stress as key determinants of CVD. Many

of these pathways, lipid metabolism in particular, have shown marked differences between men and women, which may in part explain the sex differences observed in CVD risk. For example, VLDL triglyceride content is higher in men relative to women, while inflammatory glycoprotein acetyl is higher in women across different ages.³⁴

Gut microbiome

Metabolomics can capture metabolic markers of gut microbiota, which have been increasingly recognised to have an important role in cardiometabolic health. Particularly, trimethylamine-N-oxide (TMAO), a gut microbial-host cometabolite of dietary choline and carnitine, was observed to be strongly associated with CHD and diabetes.^{3,4,35} Supplementation with TMAO or its precursors promoted atherosclerosis in mice, but this effect was not observed in germ-free mice indicating that a gut-microbial step was involved. Specifically TMAO involves the gut microbial transformation of dietary PC and carnitine,³⁶ typical components of a meat-based diet which, in turn, is associated with poor cardiometabolic health (figure 4). Animal studies have shown that increasing TMAO promotes thrombosis, atherosclerosis and metabolic dysfunction, whereas inhibiting TMAO reduces the formation of atherosclerotic lesions and thrombosis potential and improves glucose tolerance.^{37–39}

Multionics analyses and MR

Several studies have integrated genetic and metabolomics data using the MR approach to explore potential causality. In MR, genetic variants associated with metabolite levels are used

Figure 4 TMAO pathway and its association with cardiovascular disease. TMA, trimethylamine; TMAO, trimethylamine-N-oxide.

as instrumental variables to investigate their effects on CVD (figure 5). Since genetic variants are randomly assigned during meiosis, using genetic variants as instruments for metabolite levels mimics the design of a randomised controlled trial.⁴⁰ For example, MR analyses have supported a possible causal role of MG 18:2 and apolipoprotein B on CHD risk.⁴¹ However, MR analysis relies on specific assumptions such as the absence of horizontal pleiotropy, that is, the effects of the genetic variants on the outcome are assumed to act exclusively through the risk factor (metabolite) of interest. In relation to metabolomics, many lipid-related genetic variants are highly pleiotropic and therefore the standard MR paradigm may not be suitable.⁴² Systems biology approaches may also be used to highlight links between biological pathways. For example, integration of gene and metabolite networks revealed an interaction network between genes mainly involved in inflammatory, insulin and lipid pathways, and NMR-identified metabolite markers of atherosclerosis (figure 3).⁵

Clinical utility and risk stratification

As well as indicating pathways underpinning CVD, metabolites may serve as novel disease biomarkers to aid CVD risk stratification for disease prevention, although often these studies have been small and underpowered, and their incremental predictive ability over and above traditional risk factors has been limited.^{5 8 27 30 31} Given the importance of lipids in CVD risk stratification, research effort has concentrated on the role of

lipid species and, in particular, in lipoprotein particles. Among these, only apolipoprotein-B and lipoprotein A have shown some evidence for clinical utility in specific patient populations, whereas the clinical utility of lipoprotein subclasses and apolipoprotein profiles remains to be established.^{43 44}

CONCLUSIONS AND FUTURE DIRECTIONS

Metabolomics presents a powerful and promising approach for an in-depth molecular characterisation of cardiovascular health and disease. Advances in analytical technologies and informatics have made possible the generation and analysis of high-dimensional complex metabolic phenotyping data in large population studies. Such studies have highlighted the systemic nature of CVD, including the role of gut microbial cometabolism, branched chain amino acids, glycerophospholipid and cholesterol metabolism, as well as activation of inflammatory processes. Although lipid and inflammatory pathways are already well known with respect to CVD, metabolic phenotyping produces a higher-resolution and comprehensive biological signature, which may lead to new understanding of mechanisms and novel drug targets. The clinical utility of metabolites in risk stratification to predict future cardiovascular risk among healthy individuals, over and above current CVD risk algorithms, has not yet been established. Efforts to harmonise and synthesise the differently generated metabolomic data across studies and platforms are needed to enhance the generalisability, reproducibility, molecular scope and potential for clinical utility of this approach. In addition, integration of metabolomics data with other orthogonal technologies such as proteomics, genomics and epigenetics will provide an even deeper understanding of the underlying biological pathways and mechanisms. In this regard, validation and follow-up of epidemiological findings with mechanistic investigations such as in vitro and animal studies would strengthen the evidence for cause-and-effect relationships.

Contributors IT and AI drafted the paper; PE supervised the project; and all authors made critical revisions.

Funding PE is director of the MRC Centre for Environment and Health funded by the UK Medical Research Council (MR/L01341X/1, MR/S019669/1). PE and IT are supported by the British Heart Foundation Centre for Research Excellence at Imperial College London and the National Institute for Health Research Imperial Biomedical

Figure 5 Example of Mendelian randomisation concept applied to metabolites.

Research Centre, Imperial College London. AI is supported by a Stavros Niarchos Foundation grant.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iD

Paul Elliott <http://orcid.org/0000-0002-7511-5684>

REFERENCES

- Ezzati M, Obermeyer Z, Tzoulaki I, et al. Contributions of risk factors and medical care to cardiovascular mortality trends. *Nat Rev Cardiol* 2015;12:508–30.
- Tzoulaki I, Elliott P, Kontis V, et al. Worldwide exposures to cardiovascular risk factors and associated health effects: current knowledge and data gaps. *Circulation* 2016;133:2314–33.
- Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 2011;472:57–63.
- Tang WHW, Wang Z, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 2013;368:1575–84.
- Tzoulaki I, Castagné R, Boulangé CL, et al. Serum metabolic signatures of coronary and carotid atherosclerosis and subsequent cardiovascular disease. *Eur Heart J* 2019;40:2883–96.
- Würtz P, Wang Q, Soininen P, et al. Metabolomic profiling of statin use and genetic inhibition of HMG-CoA reductase. *J Am Coll Cardiol* 2016;67:1200–10.
- Shah SH, Sun J-L, Stevens RD, et al. Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. *Am Heart J* 2012;163:844–50.
- Ward-Caviness CK, Xu T, Aspelund T, et al. Improvement of myocardial infarction risk prediction via inflammation-associated metabolite biomarkers. *Heart* 2017;103:1278–85.
- Ganna A, Salihovic S, Sundström J, et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS Genet* 2014;10:e1004801.
- Köttgen A, Raffler J, Sekula P, et al. Genome-wide association studies of metabolite concentrations (mGWAS): relevance for nephrology. *Semin Nephrol* 2018;38:151–74.
- García-Pérez I, Posma JM, Serrano-Contreras JI, et al. Identifying unknown metabolites using NMR-based metabolic profiling techniques. *Nat Protoc* 2020;15:2538–67.
- Lewis GD, Wei R, Liu E, et al. Metabolite profiling of blood from individuals undergoing planned myocardial infarction reveals early markers of myocardial injury. *J Clin Invest* 2008;118:3503–12.
- Karaman I. Preprocessing and pretreatment of metabolomics data for statistical analysis. *Adv Exp Med Biol* 2017;965:145–61.
- Dunn WB, Broadhurst D, Begley P, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2011;6:1060–83.
- Tzoulaki I, Ebbels TMD, Valdes A, et al. Design and analysis of metabolomics studies in epidemiologic research: a primer on -omic technologies. *Am J Epidemiol* 2014;180:129–39.
- Castagné R, Boulangé CL, Karaman I, et al. Improving Visualization and Interpretation of Metabolome-Wide Association Studies: An Application in a Population-Based Cohort Using Untargeted ¹H NMR Metabolic Profiling. *J Proteome Res* 2017;16:3623–33.
- Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis. *Metabolomics* 2007;3:211–21.
- Posma JM, Robinette SL, Holmes E, et al. MetaboNetworks, an interactive Matlab-based toolbox for creating, customizing and exploring sub-networks from KEGG. *Bioinformatics* 2014;30:893–5.
- Chong J, Soufan O, Li C, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* 2018;46:W486–94.
- Yu B, Zanetti KA, Tempora M, et al. The Consortium of metabolomics studies (COMETS): metabolomics in 47 prospective cohort studies. *Am J Epidemiol* 2019;188:991–1012.
- Jayasinghe SU, Tankeu AT, Amati F. Reassessing the role of diacylglycerols in insulin resistance. *Trends Endocrinol Metab* 2019;30:618–35.
- Lind L, Salihovic S, Ganna A, et al. A multi-cohort metabolomics analysis discloses sphingomyelin (32:1) levels to be inversely related to incident ischemic stroke. *J Stroke Cerebrovasc Dis* 2020;29:104476.
- Edsfeldt A, Dunér P, Ståhlman M, et al. Sphingolipids contribute to human atherosclerotic plaque inflammation. *Arterioscler Thromb Vasc Biol* 2016;36:1132–40.
- Li Y, Talbot CL, Chaurasia B. Ceramides in adipose tissue. *Front Endocrinol* 2020;11.
- Summers SA. Could ceramides become the new cholesterol? *Cell Metab* 2018;27:276–80.
- Seah JYH, Chew WS, Torta F, et al. Plasma sphingolipids and risk of cardiovascular diseases: a large-scale lipidomic analysis. *Metabolomics* 2020;16:89.
- Stegemann C, Pechlaner R, Willeit P, et al. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. *Circulation* 2014;129:1821–31.
- Sanders FWB, Acharjee A, Walker C, et al. Hepatic steatosis risk is partly driven by increased de novo lipogenesis following carbohydrate consumption. *Genome Biol* 2018;19:79.
- Holmes MV, Millwood IY, Kartsonaki C, et al. Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. *J Am Coll Cardiol* 2018;71:620–32.
- Würtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk. *Circulation* 2015;131:774–85.
- Vaahoranta AAM, Verhoeven A, Weller CM, et al. A metabolomic profile is associated with the risk of incident coronary heart disease. *Am Heart J* 2014;168:45–52.
- White PJ, Newgard CB. Branched-chain amino acids in disease. *Science* 2019;363:582–3.
- Wittebans LBL, Lotta LA, Oliver-Williams C, et al. Assessing the causal association of glycine with risk of cardio-metabolic diseases. *Nat Commun* 2019;10:1060.
- Bell JA et al. Sex differences in cardiometabolic traits at four life stages: cohort study with repeated metabolomics. *medRxiv* 2020.
- Heianza Y, Ma W, Manson JE, et al. Gut microbiota metabolites and risk of major adverse cardiovascular disease events and death: a systematic review and meta-analysis of prospective studies. *J Am Heart Assoc* 2017;6. doi:10.1161/JAHA.116.004947. [Epub ahead of print: 29 Jun 2017].
- Koeth RA, Wang Z, Levison BS, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med* 2013;19:576–85.
- Chen S, Henderson A, Petriello MC, et al. Trimethylamine N-oxide binds and activates PERK to promote metabolic dysfunction. *Cell Metab* 2019;30:1141–51.
- Roberts AB, Gu X, Buffa JA, et al. Development of a gut microbe-targeted nonlethal therapeutic to inhibit thrombosis potential. *Nat Med* 2018;24:1407–17.
- Wang Z, Roberts AB, Buffa JA, et al. Non-lethal inhibition of gut microbial trimethylamine production for the treatment of atherosclerosis. *Cell* 2015;163:1585–95.
- von Hinke S, Davey Smith G, Lawlor DA, et al. Genetic markers as instrumental variables. *J Health Econ* 2016;45:131–48.
- Zuber V, Gill D, Ala-Korpela M, et al. High-throughput multivariable Mendelian randomization analysis prioritizes apolipoprotein B as key lipid risk factor for coronary artery disease. *Int J Epidemiol* 2020. doi:10.1093/ije/dyaa216. [Epub ahead of print: 01 Nov 2020].
- Zuber V, Colijn JM, Klaver C, et al. Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization. *Nat Commun* 2020;11:29.
- Nordestgaard BG, Langlois MR, Langsted A, et al. Quantifying atherogenic lipoproteins for lipid-lowering strategies: consensus-based recommendations from EAS and EFLM. *Atherosclerosis* 2020;294:46–61.
- Jellinger PS, Smith DA, Mehta AE, et al. American association of clinical endocrinologists' guidelines for management of dyslipidemia and prevention of atherosclerosis. *Endocr Pract* 2012;18 Suppl 1:1–78.