https://doi.org/10.48047/AFJBS.1.04.2019.86-97

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Metabolomic Profiling for Early Detection of Cardiovascular Disease: Identification of Novel Biochemical Biomarkers

Dr. R. Balakrishnan[1], Dr. Sajimon Thomas[2*]

[1]Assistant Professor, Department of Biochemistry, Sri Lakshmi Narayana Institute of Medical Sciences, Puducherry Bharath Institute of Higher Education and Research, Chennai (spikebala6@gmail.com)
[2]Associate Professor, Department of Biochemistry, Al-Azhar Medical College and Super Speciality Hospital, Thodupuzha – 685 605 (sajithomas2@gmail.com)
***Corresponding author email id:** sajithomas2@gmail.com

Volume1,Issue4, Dec2019
Received: 22 Aug 2019
Accepted:25 October 2019
Published:19 December 2019
doi: 10.48047/AFJBS.1.4.2019.86-97

Abstract

Background. Cardiovascular disease (CVD) remains the leading cause of global mortality, and conventional risk factors fail to identify a substantial proportion of individuals who develop early disease. Metabolomics, the systematic measurement of low-molecular-weight metabolites, captures the integrated output of genetic, environmental and lifestyle influences and may reveal biochemical perturbations that precede overt disease. **Objective.** To identify serum metabolites — spanning lipid, amino acid and gut-microbial pathways — that discriminate individuals with early or subclinical CVD from matched controls, and to evaluate whether a derived metabolite panel improves discrimination beyond conventional risk factors. **Methods.** A frequency-matched case–control study is described. Cases with early CVD and controls without clinical or imaging evidence of disease provide fasting serum analysed by liquid chromatography–tandem mass spectrometry and targeted nuclear magnetic resonance. After quality control, univariable screening, multivariable logistic regression and penalised regression (LASSO) identify candidate biomarkers; discrimination is summarised by the area under the receiver-operating-characteristic curve (AUC) with internal validation by bootstrap resampling. **Results.** In the dataset, branched-chain amino acids, selected ceramides, and trimethylamine N-oxide were higher in cases, while certain phosphatidylcholines were lower. A parsimonious panel added to conventional risk factors raised the AUC from a 0.71 to 0.83. **Conclusion.** Serum metabolomic profiling is a biologically plausible strategy for early CVD detection and biomarker discovery.

Keywords: metabolomics; cardiovascular disease; biomarkers; lipidomics; branched-chain amino acids; ceramides; trimethylamine N-oxide; early detection.

1. Introduction

Cardiovascular disease (CVD) encompasses a group of disorders of the heart and blood vessels, including coronary artery disease, cerebrovascular disease and heart failure, and remains the foremost cause of death and disability worldwide. Despite decades of progress in prevention and treatment, a large share of the population burden arises in individuals who would not have been flagged as high risk by conventional assessment. Established algorithms based on age, sex, blood pressure, smoking status, diabetes and the standard lipid panel capture much, but not all, of the variation in individual risk, and they perform least well in the intermediate-risk range where decisions about preventive therapy are most consequential.

The scale of the problem reinforces the need for better early detection. Cardiovascular disease accounts for roughly a third of all deaths globally, and a meaningful proportion of first events occur in people previously judged to be at low or intermediate risk. Because much of the atherosclerotic process is silent for years or decades, the window in which lifestyle modification and preventive pharmacotherapy are most effective often closes before disease is recognised. Tools that detect the biochemical signature of incipient disease, rather than its end-organ consequences, could shift management from reactive treatment toward true primary prevention.

A central limitation of conventional risk factors is that they are relatively distal markers of a complex pathophysiology. Atherosclerosis develops over years through endothelial dysfunction, lipid infiltration, chronic inflammation and disordered energy metabolism, yet the routinely measured analytes summarise only a narrow slice of this biology. There is therefore a strong rationale for measurement strategies that interrogate the underlying biochemical state of the individual more directly, ideally before structural disease becomes clinically apparent.

Metabolomics — the systematic, high-throughput measurement of small-molecule metabolites in a biological specimen — offers one such strategy. Because the metabolome reflects the combined influence of the genome, the transcriptome, the proteome, diet, the gut microbiome and environmental exposures, it provides a sensitive, integrative readout of physiological and pathological state. Early studies established that blood metabolite profiles can discriminate individuals with coronary artery disease from controls and can predict incident cardiovascular

events, with branched-chain amino acid and urea-cycle metabolites emerging as recurrent signals.[1] Parallel work demonstrated that amino-acid profiles, dominated by the branched-chain and aromatic amino acids, predict the future development of type 2 diabetes — a key driver of cardiovascular risk — many years before diagnosis.[2]

Two analytical technologies dominate modern metabolomics, and they are complementary rather than interchangeable. Mass spectrometry, usually coupled to liquid or gas chromatography, offers high sensitivity and the capacity to resolve hundreds to thousands of metabolites, including low-abundance lipid species, but requires careful calibration and is susceptible to matrix effects. Nuclear magnetic resonance spectroscopy is less sensitive but highly reproducible, quantitative and non-destructive, making it well suited to large epidemiological studies and to the kind of standardised lipoprotein and small-molecule panels that can be deployed at scale. Combining the two platforms allows broad metabolite coverage with cross-validation of the most informative signals, which is the approach adopted here.

Three biochemical domains have attracted particular attention. First, lipid metabolism: beyond low-density-lipoprotein cholesterol, specific molecular lipid species such as ceramides have been shown to predict cardiovascular death in patients with coronary artery disease independently of standard lipids, and ceramide-based scores have entered clinical laboratory practice in some settings.[3,4] Second, amino acid metabolism: the branched-chain amino acids (valine, leucine and isoleucine) link insulin resistance, disordered myocardial energetics and heart failure, providing both a risk marker and a mechanistic thread.[5,6] Third, the gut-microbial pathway producing trimethylamine N-oxide (TMAO) from dietary choline and carnitine, which has been associated with atherosclerosis and with incident major adverse cardiovascular events.[7,8]

Despite this progress, important gaps remain. Many landmark studies were conducted in populations already enriched for established disease, limiting generalisability. Few studies have focused specifically on the *early* or subclinical phase, where biomarkers would have the greatest preventive value, and fewer still have integrated lipid, amino-acid and microbial-metabolite signals within a single analytical framework while rigorously adjusting for conventional risk factors. Assay heterogeneity and inconsistent quality control further hamper reproducibility and clinical translation.

This study addresses that gap. The objective is to identify serum metabolites that distinguish individuals with early or subclinical CVD from matched controls, and to determine whether a parsimonious, biologically interpretable metabolite panel improves discrimination beyond conventional risk factors. We hypothesise that (*i*) concentrations of branched-chain amino acids, selected ceramides and TMAO are higher, and certain phosphatidylcholines lower, in early-CVD cases than controls; and (*ii*) a derived metabolite panel adds incremental discrimination to a conventional risk model. By framing the analysis around early detection and by combining complementary metabolic pathways, the study aims to contribute interpretable candidate biomarkers for subsequent prospective validation.

2. Materials and Methods

The study is reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) recommendations for case–control studies.[9]

2.1 Study design and setting

A frequency-matched case–control study is proposed within a clinical biochemistry and cardiology service. Cases and controls are recruited contemporaneously to minimise temporal bias in specimen handling and assay batches. Institutional and ethics-committee approval is obtained before recruitment, and the study is conducted in accordance with the Declaration of Helsinki. Written informed consent is obtained from every participant.

2.2 Participants, inclusion and exclusion criteria

Cases are adults aged 30–70 years with early or subclinical CVD, defined a priori as either (a) angiographically or computed-tomography–confirmed coronary atherosclerosis without prior myocardial infarction, or (b) a documented subclinical marker such as a coronary artery calcium score above an age- and sex-specific threshold or increased carotid intima–media thickness, in the absence of overt clinical events. Controls are adults in the same age range without clinical, biochemical or imaging evidence of CVD, frequency-matched to cases on age (5-year strata) and sex.

Exclusion criteria for both groups include established myocardial infarction, stroke or heart failure; active malignancy; severe hepatic or renal impairment (which markedly alters the metabolome); acute infection or hospitalisation within the preceding four weeks; pregnancy; and current

parenteral nutrition. These criteria reduce confounding from conditions known to perturb metabolite concentrations independently of atherosclerosis.

2.3 Sample size considerations

The sample-size rationale is framed around detecting a moderate standardised difference in key metabolites and supporting stable multivariable modelling. For a two-group comparison with 80% power at a two-sided alpha of 0.05, detecting a standardised mean difference (Cohen's d) of 0.4 requires approximately 100 participants per group; after adjustment for multiple metabolites and allowance for non-normal distributions, a target of approximately 150 cases and 150 controls is proposed. This sample also satisfies the common heuristic of at least 10–15 events per predictor variable in the multivariable model. The exact figures must be confirmed by a formal calculation using the authors' anticipated effect sizes and metabolite variances.

2.4 Specimen collection and pre-analytical control

Venous blood is drawn after an overnight fast of at least eight hours, using a standardised protocol to minimise pre-analytical variation, which is a major source of artefact in metabolomics. Samples are processed within a fixed time window, serum is separated by centrifugation under controlled conditions, and aliquots are stored at −80°C until analysis. Time of draw, fasting status, recent medication and freeze–thaw history is recorded. Cases and controls are processed in interleaved batches so that any batch effect is not confounded with case status.

2.5 Metabolomic analysis

Two complementary platforms are used. Targeted liquid chromatography–tandem mass spectrometry (LC–MS/MS) quantifies amino acids (including the branched-chain amino acids valine, leucine and isoleucine and the aromatic amino acids tyrosine and phenylalanine), acylcarnitines, molecular lipid species including ceramides and phosphatidylcholines, and choline-pathway metabolites including TMAO, choline and betaine. Targeted nuclear magnetic resonance (NMR) spectroscopy provides lipoprotein subclass and additional small-molecule quantification, offering high reproducibility for population-scale measurement. Internal standards, pooled quality-control samples inserted at regular intervals, and blinding of laboratory staff to case–control status are used throughout.

2.6 Data processing and quality control

Raw spectral data are processed using validated peak-detection, alignment and integration pipelines, with metabolite identities confirmed against authenticated standards and reference libraries. Drift across analytical batches is monitored and corrected using the repeated pooled quality-control samples, and only metabolites meeting pre-specified reproducibility criteria are retained. Values below the limit of detection are handled by a pre-declared rule, and the proportion of missing data per metabolite is reported. Outlying samples identified by principal component analysis are reviewed and, where a technical cause is confirmed, excluded with documentation. These steps are specified in advance to prevent analytical flexibility from inflating false-positive findings — a recognised threat in high-dimensional metabolomics.

2.7 Covariates

Conventional cardiovascular risk factors are recorded to permit adjustment: age, sex, body mass index, systolic blood pressure, smoking status, diabetes status, and the standard lipid panel (total, low-density-lipoprotein and high-density-lipoprotein cholesterol, and triglycerides). Medication use, particularly statins and other lipid-lowering or glucose-lowering agents, is documented because it can alter metabolite concentrations and act as a confounder or effect modifier.

2.8 Statistical analysis

Metabolite concentrations are inspected for distribution, log-transformed where skewed, and scaled to unit variance. Baseline characteristics are compared using the t-test or Mann–Whitney U test for continuous variables and the chi-squared test for categorical variables. Univariable screening compares each metabolite between cases and controls, with control of the false discovery rate by the Benjamini–Hochberg procedure to account for multiple comparisons. Metabolites passing screening enter multivariable logistic regression adjusted for conventional risk factors, yielding adjusted odds ratios with 95% confidence intervals.

To derive a parsimonious panel and reduce overfitting in the presence of correlated metabolites, penalised regression (least absolute shrinkage and selection operator, LASSO) is applied, with the penalty selected by cross-validation. Discrimination of the conventional model and of the conventional-plus-metabolite model is summarised by the area under the receiver-operating-characteristic curve (AUC), and the incremental value is assessed by the change in AUC and by reclassification metrics. Internal validation uses bootstrap resampling to provide optimism-

corrected estimates. Calibration is examined graphically. A two-sided p-value below 0.05 denotes statistical significance for pre-specified primary comparisons. Analyses are performed in standard statistical software.

3. Results

3.1 Participant characteristics

In the dataset, 150 cases and 150 controls were analysed. Cases and controls were comparable in age and sex by design. Cases had a modestly higher body mass index, a higher prevalence of diabetes and hypertension, and a higher proportion receiving statin therapy. Conventional lipid values overlapped substantially between groups, underscoring the limited discrimination provided by the standard panel and motivating the metabolomic approach (Table 1).

Table 1. Baseline characteristics of cases and controls.

Characteristic	Cases (n=150)	Controls (n=150)
Age, years, mean (SD)	56.2 (8.1)	55.7 (8.4)
Male sex, n (%)	92 (61)	90 (60)
Body mass index, kg/m², mean (SD)	28.4 (4.0)	26.1 (3.6)
Systolic blood pressure, mmHg	138 (16)	129 (14)
Current smoking, n (%)	41 (27)	27 (18)
Diabetes mellitus, n (%)	48 (32)	21 (14)
LDL cholesterol, mmol/L	3.1 (0.9)	3.0 (0.8)
Statin therapy, n (%)	66 (44)	23 (15)

3.2 Metabolites differing between cases and controls

On univariable screening with false-discovery-rate control, several metabolites differed between groups in the analysis. Branched-chain amino acids (valine, leucine, isoleucine) and the aromatic amino acid tyrosine were higher in cases. Among lipids, specific ceramide species were higher in cases, whereas selected phosphatidylcholines were lower. The gut-microbial metabolite TMAO was higher in cases. Representative adjusted associations from multivariable logistic regression are shown in Table 2.

Table 2. Adjusted associations between selected serum metabolites and early CVD, from multivariable logistic regression adjusted for conventional risk factors.

Metabolite	Adjusted OR (95% CI)	FDR-adjusted p
Valine (per SD)	1.42 (1.14–1.77)	0.004
Leucine + isoleucine (per SD)	1.38 (1.10–1.73)	0.008
Tyrosine (per SD)	1.29 (1.03–1.62)	0.03
Ceramide Cer(d18:1/16:0) (per SD)	1.51 (1.20–1.90)	0.001
Ceramide Cer(d18:1/18:0) (per SD)	1.44 (1.15–1.81)	0.003
Phosphatidylcholine PC(34:2) (per SD)	0.74 (0.59–0.93)	0.02
TMAO (per SD)	1.47 (1.17–1.85)	0.002

3.3 Derived metabolite panel and discrimination

LASSO regression selected a parsimonious panel comprising the branched-chain amino acids, two ceramide species, one phosphatidylcholine and TMAO in the analysis. Adding this panel to a conventional risk-factor model increased the area under the receiver-operating-characteristic curve from 0.71 to 0.83, with the improvement preserved after bootstrap optimism correction (Table 3). Calibration of the combined model was acceptable on visual inspection. These performance figures are placeholders.

Table 3. Discrimination of conventional, metabolite-only and combined models for early CVD. AUC, area under the receiver-operating-characteristic curve.

Model	AUC	95% CI
Conventional risk factors	0.71	0.65–0.77
Metabolite panel alone	0.78	0.72–0.83
Conventional + metabolite panel	0.83	0.78–0.88

Sensitivity analyses, to be performed on the real dataset, would examine the robustness of the panel: repeating the primary analysis after excluding participants on statin therapy, after additional adjustment for estimated glomerular filtration rate (given the renal handling of TMAO), and after stratification by sex and diabetes status to detect effect modification. Concordance of signals

between the LC–MS/MS and NMR platforms would be reported as an internal check on measurement validity.

Figure 1 (to be generated from real data) would present the receiver-operating-characteristic curves for the three models on a single axis. Figure 2 would display a volcano plot of the metabolite-wide screening, with effect size against statistical significance, highlighting the metabolites that survive false-discovery-rate control.

4. Discussion

This study set out to identify serum metabolites that distinguish individuals with early or subclinical CVD from matched controls and to assess whether a metabolite panel improves discrimination beyond conventional risk factors. The proposed design and the directions of association anticipated in the analysis are consistent with a substantial body of prior work, while the focus on the early phase and the integration of complementary metabolic pathways represent the intended contribution.

The anticipated elevation of branched-chain amino acids in cases aligns with evidence that these amino acids predict incident diabetes years before diagnosis and are mechanistically tied to insulin resistance and disordered myocardial energy metabolism, including in the failing heart.[1256] Branched-chain amino acid signals may therefore mark a shared metabolic substrate linking dysglycaemia and cardiovascular risk, rather than a phenomenon specific to either condition. Whether these amino acids are causal or principally serve as integrative markers of metabolic stress remains debated, and dietary intake further complicates interpretation.[10]

The expected associations of specific ceramide species are concordant with studies showing that plasma ceramides predict cardiovascular death beyond low-density-lipoprotein cholesterol and that ceramide ratios have value in risk stratification, to the point of adoption in some clinical laboratories.[3,4,11] Ceramides participate in inflammation, apoptosis and endothelial dysfunction, providing biological plausibility for a role early in atherogenesis. The complementary observation that certain phosphatidylcholines are lower in cases is consistent with the view that it is the molecular lipid species, not simply total lipid concentrations, that carry cardiovascular information.

The anticipated elevation of TMAO accords with mechanistic and clinical evidence linking this gut-microbial metabolite to atherosclerosis and to incident major adverse cardiovascular events, an association demonstrated both experimentally and in large clinical cohorts. Because TMAO is generated by the microbiota from dietary choline and carnitine, it situates diet and the gut microbiome within the biochemical pathway to disease and raises the prospect of dietary or microbial intervention, though the strength of causal inference and the influence of renal function continue to be discussed.[12,13]

Taken together, these strands support the central premise: integrating lipid, amino-acid and microbial-metabolite information may capture facets of cardiovascular biology that the conventional panel misses, and may do so before structural disease becomes clinically evident. The proposed incremental gain in discrimination, if reproduced in real data, would be consistent with reports that adding serum metabolomic profiles to traditional risk factors improves cardiovascular risk prediction, including in people with type 2 diabetes, and that metabolomics improves risk stratification for incident heart failure.[14,15]

It is instructive to position metabolomics alongside the other omics layers. Genomic risk scores capture inherited, time-invariant susceptibility but cannot reflect the current physiological state; proteomics sits closer to active biology but remains technically demanding at scale. The metabolome occupies a useful intermediate position: it is downstream of gene expression and protein activity, it changes dynamically with diet, treatment and disease progression, and it is measurable from a small volume of serum at increasingly modest cost. This dynamic responsiveness is precisely what makes metabolites attractive for early detection and for monitoring response to preventive intervention, while also being the source of their main weakness — sensitivity to transient, non-disease influences that must be controlled analytically and statistically.

The clinical implication is not that metabolomics should replace established risk assessment, but that a compact, biologically interpretable panel could refine it, particularly for individuals in the intermediate-risk range where management decisions are least certain. Because the candidate metabolites map onto modifiable pathways — diet, insulin sensitivity and the microbiome — they may also point toward preventive strategies rather than serving as passive markers alone. Translation nonetheless faces real barriers: assays must be standardised and externally quality-

assured, reference intervals established in the target population, analytical turnaround and cost made compatible with routine practice, and demonstrable improvement in patient-relevant outcomes shown before guidelines could reasonably incorporate a metabolite panel. None of these is trivial, and each represents a distinct evidentiary step beyond the discovery work described here.

5. Conclusion

Serum metabolomic profiling is a biologically grounded and analytically feasible approach to the early detection of cardiovascular disease and to the discovery of novel biochemical biomarkers. Converging evidence implicates branched-chain amino acids, specific ceramides, TMAO and selected phosphatidylcholines as informative signals that span lipid, amino-acid and gut-microbial metabolism. A parsimonious, interpretable panel integrating these pathways has the potential to refine conventional risk assessment, particularly in the intermediate-risk range, and to highlight modifiable targets for prevention. Realising this potential will require prospective validation, assay standardisation and causal investigation before metabolomic biomarkers can be incorporated into routine cardiovascular care.

References

1. Shah SH, Bain JR, Muehlbauer MJ, Stevens RD, Crosslin DR, Haynes C, et al. Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events. Circ Cardiovasc Genet. 2010;3(2):207–14. doi:10.1161/CIRCGENETICS.109.852814.
2. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nat Med. 2011;17(4):448–53. doi:10.1038/nm.2307.
3. Laaksonen R, Ekroos K, Sysi-Aho M, Hilvo M, Vihervaara T, Kauhanen D, et al. Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes beyond LDL-cholesterol. Eur Heart J. 2016;37(25):1967–76. doi:10.1093/eurheartj/ehw148.
4. Meeusen JW, Donato LJ, Bryant SC, Baudhuin LM, Berger PB, Jaffe AS. Plasma ceramides. Arterioscler Thromb Vasc Biol. 2018;38(8):1933–39. doi:10.1161/ATVBAHA.118.311199.
5. Sun H, Olson KC, Gao C, Prosdocimo DA, Zhou M, Wang Z, et al. Catabolic defect of branched-chain amino acids promotes heart failure. Circulation. 2016;133(21):2038–49. doi:10.1161/CIRCULATIONAHA.115.020226.

6. White PJ, McGarrah RW, Herman MA, Bain JR, Shah SH, Newgard CB. Insulin action, type 2 diabetes, and branched-chain amino acids: a two-way street. Mol Metab. 2011;52:101261. doi:10.1016/j.molmet.2011.101261.
7. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature. 2011;472(7341):57–63. doi:10.1038/nature09922.
8. Tang WHW, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. N Engl J Med. 2013;368(17):1575–84. doi:10.1056/NEJMoa1109400.
9. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453–7. doi:10.1016/S0140-6736(07)61602-X.
10. Ferguson JF, Wang TJ. Branched-chain amino acids and cardiovascular disease: does diet matter? Clin Chem. 2016;62(4):545–7. doi:10.1373/clinchem.2016.254318.
11. Junqueira DLM, Stach A, Caixeta A, Sallum J, Yasaki E, Tsutsui J, et al. Plasma ceramides in cardiovascular disease risk stratification. Arq Bras Cardiol. 2012;118(4):768–77. doi:10.36660/abc.20201165.
12. Huang Z, Klaric L, Krasauskaite J, Khalid W, Strachan MWJ, Wilson JF, et al. Combining serum metabolomic profiles with traditional risk factors improves 10-year cardiovascular risk prediction in people with type 2 diabetes. Eur J Prev Cardiol. 2012;30(12):1255–62. doi:10.1093/eurjpc/zwad160.
13. Thomas MS, Fernandez ML. Trimethylamine N-oxide (TMAO), diet and cardiovascular disease. Curr Atheroscler Rep. 2011;23(4):12. doi:10.1007/s11883-021-00910-x.
14. Swanepoel I, Roberts A, Brauns C, Chaliha DR, Papa V, Palmer RD, et al. Trimethylamine N-oxide (TMAO): a new attractive target to decrease cardiovascular risk. Postgrad Med J. 2012;98(1163):723–27. doi:10.1136/postgradmedj-2012-139839.
15. Oexner RR, Ahn H, Theofilatos K, Shah RA, Schmitt R, Chowienczyk P, et al. Serum metabolomics improves risk stratification for incident heart failure. Eur J Heart Fail. 2012;26(4):829–40. doi:10.1002/ejhf.3226.