Untargeted Metabolomics in the Discovery of Novel Biomarkers and Therapeutic Targets for Atherosclerotic Cardiovascular Diseases

Vi T. Dang\textsuperscript{1,2}, Aric Huang\textsuperscript{1,2} and Geoff H. Werstuck\textsuperscript{1,2,3,*}

\textsuperscript{1}Thrombosis and Atherosclerosis Research Institute, \textsuperscript{2}Department of Chemistry and Chemical Biology, \textsuperscript{3}Department of Medicine, McMaster University, Hamilton, Ontario, Canada

Abstract: Background: Cardiovascular Disease (CVD) is the leading cause of mortality and morbidity worldwide. Four out of five CVD deaths are due to myocardial infarction or stroke. Despite many initiatives that have been established for CVD prevention and risk management, and new therapies to treat existing CVD, patients continue to die from cardiac events. Clearly, we need to identify new therapeutic targets and strategies. Metabolomics offers a novel solution to this problem, as metabolomics-based biomarkers do not only indicate the presence or absence of a disease, but are also capable of assessing risks of developing the disease and detecting the disease prior to the appearance of overt clinical symptoms.

Method: In this review, we describe the analytical techniques and workflow used in untargeted metabolomics. We also identify several case studies that highlight the use of untargeted metabolomics in cardiovascular research.

Results: Five case studies that employ untargeted metabolomics approaches to identify biomarkers for cardiovascular risk, myocardial ischemia, transient ischemic attack, incident coronary heart disease, and myocardial infarction risk prediction are described. The use of the untargeted metabolomics is still relatively new in cardiovascular research. As such, there remains a need for future advancement in metabolomic technologies.

Conclusion: Early diagnosis of CVDs and identification of patients at high risk of developing adverse events would allow for timely intervention that prevents serious consequences or death. There is a need to establish sensitive and non-invasive CV biomarkers, and novel therapeutic targets for the prevention and treatment of CVDs.

Keywords: Metabolomics, cardiovascular disease, biomarker, myocardial infarction, coronary heart disease, atherosclerotic.

1. INTRODUCTION: CARDIOVASCULAR DISEASE

Cardiovascular Disease (CVD) is one of the major global health concerns of our century \cite{1}. Together with cancer, chronic respiratory diseases and diabetes, CVD is one of four priority non-communicable diseases targeted for action by the World Health Organization. CVD is responsible for approximately 31\% of all deaths worldwide, which makes it the leading cause of mortality \cite{2, 3}. In 2013, 17.3 million deaths were attributed to CVDs \cite{4}. This number is expected to increase to over 23.6 million deaths by 2030. CVDs impose a heavy burden on our healthcare systems. The estimated global cost of CVD was $863 billion in 2010 and this is expected to rise to $1044 billion by 2030 \cite{4}.

Several different independent risk factors have been linked to the development of CVDs. Some of these risk factors can be altered and/or treated, while other risk factors cannot be controlled. Modifiable risk factors include hypertension (elevated blood pressure, resting blood pressure >140/90 mmHg), dyslipidemia (total cholesterol >240 mg/dL, low-density lipoprotein >160 mg/dL and high density lipoprotein <40 mg/dL), physical inactivity, obesity (body mass index >30 kg/m\textsuperscript{2}), diabetes mellitus (elevated blood glucose, fasting glucose level >140 mg/dL), and tobacco use \cite{2, 5}. Non-modifiable risk factors include advancing age, ethnic characteristics (African Americans, Hispanics, and non-Hispanic whites are at higher risk), and family history (genetics).

Eight risk factors (alcohol use, tobacco use, high blood glucose, high blood pressure, high body mass index, high cholesterol, low fruit and vegetable intake, and physical inactivity) account for 61\% of cardiovascular deaths \cite{6}.
Combined, these same risk factors account for over three-quarters of ischaemic heart disease. Reducing exposure to these eight risk factors would increase global life expectancy by almost 5 years.

2. ATHEROSCLEROSIS

Four out of five CVD deaths are due to Myocardial Infarction (MI) and stroke [6]. The underlying cause of these diseases is atherosclerosis, which is a multifactorial, progressive and chronic inflammatory disease of the medium-large arteries (Fig. 1). Atherosclerosis is characterized by the accumulation of lipids, inflammatory factors and fibrous elements within the arterial wall, resulting in the hardening and narrowing of blood vessels [7].

Atherosclerotic plaques, also known as lesions, initially accumulate in regions with turbulent, non-laminar blood flow, including areas with bifurcations, branches, or inner curvatures [8]. At these sites, there is an increased expression of endothelial cell surface proteins, including vascular cell adhesion molecule-1 and P-selectin, which mediates the migration of monocytes and T lymphocytes across the endothelium into the sub-endothelial intima layer [9, 10]. Once in the sub-endothelial space, monocytes differentiate into macrophages that endocytose low-density lipoprotein (LDL) and modified-LDL particles. The lipid engorged macrophages, known as foam cells, make up the fatty streak which is the earliest type of lesion.

Fatty streaks are benign but can mature into more advanced lesions. Macrophage foam cells and lymphocytes amplify the inflammatory response by secreting cytokines and growth factors, including interferon-γ, interleukin-1, and tumor necrosis factor-α [11, 12]. In addition, cytokines induce the migration of vascular smooth muscle cells from the tunica media into the intima [13]. Advanced lesions are characterized by a lesion-stabilizing fibrous cap, containing vascular smooth muscle cells and their synthesized collagen, which covers the lipid core. As the lesion grows, foam cells continue to uptake lipids in the form of LDL and modified-LDLs. The accumulation of free cholesterol in the foam cells can ultimately initiate apoptosis. Apoptotic cells are typically cleared by macrophages through efferocytosis. However, efferocytosis is defective in advanced lesions, resulting in the formation of an acellular region within the lesion known as the necrotic core.

Necrosis is a key feature of unstable plaques that are prone to rupture. Foam cells also release collagenases, including matrix metalloproteinase-1, -2, and -9, which disrupt the biomechanical stability of the fibrous cap [14-16]. The thinning of the fibrous cap destabilizes the lesion, making it susceptible to rupture. When the lesion ruptures, coagulation factors in the blood come into contact with procoagulant factors, including tissue factor, in the necrotic core. This promotes platelet aggregation and atherothrombosis [17]. The thrombus formed may occlude the artery, resulting in
in cardiovascular complications such as myocardial ischemia or infarction.

3. THE NEED FOR NOVEL CARDIOVASCULAR BIOMARKERS AND THERAPEUTIC TARGETS

CVD mortality is expected to increase in the developing world, based on both the increasing mean population age, as well as the increase in urbanization and associated risk factor exposure [18]. Early diagnosis of CVDs and identification of patients at high risk of developing adverse cardiovascular events would allow for timely intervention to prevent serious consequences or death. Current conventional methods to identify CVD patients, or specifically patients with atherosclerosis, tend to be either invasive (e.g. angioplasty), or only have moderate predictive values (e.g. carotid ultrasound) [19-22]. Despite the existence of some successful diagnostic and prognostic biomarkers in the cardiovascular area, such as brain natriuretic peptides and troponins, many biomarkers have failed to be clinically useful because of discordant and/or heterogeneous results [23].

Recent advances have established a fundamental role of inflammation in mediating all stages of atherosclerotic progression, from the initial endothelial dysfunction to the formation and disruption of a thrombus [24-26]. Downstream biomarkers of inflammation including high sensitivity C-reactive protein and interleukin-6 were found to be associated with an increased risk of cardiovascular events, independent of cholesterol levels [27, 28]. Recently, one of the very first clinical trials on anti-inflammatory therapy for atherosclerotic disease has shown that targeting the interleukin-1β innate immunity pathway led to a significantly lower rate of recurrent cardiovascular events, independent of changes in lipid levels [29]. With these experimental and clinical findings, circulating markers of inflammation have gained substantial interest in CVD risk assessment and anti-inflammatory therapy for CVDs. Currently, inflammation markers are typically macromolecules, i.e. the interleukins.

In addition, despite many initiatives that have been established for CVD prevention and cardiovascular risk management, and new CVD therapies that have been developed in the past decade, patients continue to have cardiovascular events, suggesting that we need to identify new therapeutic targets [30]. A key factor in this mission against CVD is expanding our knowledge of the pathological molecular mechanisms underlying this complex metabolic disease. Therefore, there is an obvious need to identify sensitive and non-invasive CV biomarkers, and novel therapeutic targets for the prevention and treatment of cardiovascular events.

4. THE FIELD OF METABOLOMICS

Metabolomics is the study of metabolites which are small molecules involved in cellular metabolism. The field of metabolomics lies downstream of the central dogma of molecular biology (DNA → RNA → protein) and the metabolome (the collection of all metabolites in a biological specimen) is directly influenced by gene expression as well as environmental and dietary signals (Fig. 2). In contrast to the well-defined fields of genomics and proteomics, the study of metabolomics has only recently emerged in the late twentieth century [31, 32].

It is recognized that the levels of specific metabolites are directly linked to human health outcomes [33, 34]. In fact, the levels of metabolites are a better reflection of the functional status than the other ‘omics’ studies because the metabolic fluxes are regulated by environmental stresses in addition to gene expression [35, 36]. The genome and proteome tell us what might happen while the metabolome tells us both what might happen as well as what is currently happening in a particular tissue or organism. As such, metabolomics-based biomarkers do not only indicate the presence or absence of a disease, but are also capable of assessing the risks of developing the disease and detecting the disease prior to the appearance of overt clinical

Fig. (2). The ‘omics’ cascade. The three main fields of omics-science include genomics, proteomics, and metabolomics. Metabolomics is downstream of the central dogma of molecular biology (DNA → RNA → protein) and is the most sensitive to environmental and dietary influences. The chemical complexity of metabolomics surpasses the other ‘omics’ fields.
symptoms and/or phenotypes [37]. Currently, metabolites are used in more than 95% of all diagnostic clinical tests and approximately 90% of all known drugs are small molecules [38].

Current metabolomics technologies allow for the quantification of hundreds/thousands of circulating metabolites from multiple metabolic pathways in a single measurement. This facilitates the understanding of system-level effects of metabolites associated with a pathological condition. As such, metabolomics findings can provide insights into the molecular alterations and mechanisms underlying various physiological or pathological conditions [39, 40].

5. METABOLOMICS APPROACH

Current advances in technology allow for the simultaneous characterization of thousands of metabolites, leading to the development of so-called targeted and untargeted metabolomics [41-43]. Targeted metabolomics is commonly employed for hypothesis-driven research in which a specific list of known metabolites is measured. For example, targeted metabolomics can be used to evaluate the effect of a treatment or a surgical procedure on a specific metabolic pathway. By definition, this type of approach focuses upon a finite number of compounds/pathways of interest.

Alternatively, untargeted metabolomics, sometimes referred to as comprehensive metabolomics, is commonly used as a discovery tool in hypothesis-generating strategies that allow for the comprehensive analysis of the metabolome. With the global scope of analysis, the aim of untargeted metabolomics is to simultaneously measure as many metabolites as possible from a biological sample (e.g., cell extract, tissue or biological fluid). This allows for the examination of many enzymatic reactions and metabolic pathways at the same time. Therefore, untargeted metabolomics is capable of capturing the complexity of metabolic networks and revealing novel and unanticipated molecular alterations without bias. This type of approach can lead to the development of novel and/or unanticipated ideas about disease processes, and hence opens the doorway to new methods of diagnosis, possibly even before such conditions become clinically evident. Therefore, biomarkers discovered in untargeted metabolomics studies could potentially be used as clinical biomarkers for risk stratification and early identification of CVD.

6. ANALYTICAL TECHNOLOGIES IN UNTARGETED METABOLOMICS

The chemical complexity of metabolomics analyses far surpasses the other ‘omics’ fields (Fig. 2). Genomics must consider four nucleotide bases – adenine, cytosine, guanine and thymine, whereas proteomics must deal with twenty protein building blocks – the natural amino acids. In contrast, the metabolome is an extremely complex mixture of chemically diverse compound classes (e.g., amino acids, lipids, nucleosides, etc.), with vastly different chemical properties (e.g., hydrophilic versus hydrophobic, acid versus base) and a dynamic range of molecular concentrations (e.g., low nM to high mM) [44]. Therefore, metabolomics-based research requires cutting-edge and sophisticated analytical technologies.

With the recent developments in analytical and bioinformatic tools, metabolomics has grown exponentially in the last two decades. Nuclear Magnetic Resonance (NMR) spectroscopy and Mass Spectrometry (MS) are most commonly used in metabolomics-based research due to the high selectivity and sensitivity of these methods [45-47]. The number of PubMed publications containing the term ‘metabolomics’ and ‘mass spectrometry’ or ‘nuclear magnetic resonance’ from 2002 to 2016 are shown in Fig. 3.

NMR-based methods are widely used in metabolomics research due to their powerful capacity to provide structural information pertaining to novel analytes. However, the primary limitation of NMR spectroscopy is the lack of sensitivity. This makes it difficult to identify many biomarkers whose biological concentrations are much lower than the detection limit of NMR spectroscopy. Other drawbacks of NMR technologies are high purchasing and operating cost, low levels of throughput, and high sample volume requirement (typically in the range of low mL).

MS-based techniques are most commonly used for metabolomics research because of the very high sensitivity, selectivity, and dynamic range of MS systems. Furthermore, MS systems can be easily coupled to chromatographic separation platforms including gas chromatography, Liquid Chromatography (LC), and capillary electrophoresis. These hyphenated platforms increase the metabolite detection coverage and provide additional information (e.g. analyte retention time) for metabolite identification. Some MS instruments are capable of performing tandem (MS/MS) and

![Fig. (3). Metabolomics is a relatively new field. Recent advances in analytical and bioinformatic tools have facilitated the exponential growth of metabolomics over the last two decades. The total number of PubMed publications containing the term ‘metabolomics’ and ‘mass spectrometry’ (MS) or ‘nuclear magnetic resonance’ (NMR) from 2002 to 2016 are shown.](image-url)
multistage MS (MSn) experiments which can provide a highly desired orthogonal measurement. MS technologies also require relatively small volumes of sample (typically in the µL range). Therefore, MS platforms have been the technique of choice for many metabolomics research, especially with the untargeted approach. MS-based metabolomics publications have been growing exponentially since the mid-twentieth century.

No single analytical platform is currently capable of detecting the entire metabolome. The choice of employing a specific platform is dependent not only on the scope of the analysis and the target metabolites, but also the availability of the analytical technologies and the nature of the sample.

7. UNTARGETED METABOLOMICS WORKFLOW

The typical workflow of untargeted metabolomics consists of the following steps: metabolite extraction, data acquisition, data processing, data analysis, metabolite identification, and data interpretation.

i. Metabolite extraction: Untargeted metabolomics involves the global analysis of the metabolome. As such, sample preparation in untargeted metabolomics is optimized to extract as many metabolites as possible. An ideal extraction method should be capable of solubilizing all metabolites while excluding proteins and other higher molecular weight components. Recovery standards are commonly added prior to extraction to monitor recovery efficiency, whereas international standards are added prior to data acquisition to correct for instrumental variation.

ii. Data acquisition: No single analytical platform is currently capable of detecting the entire metabolome. Mass spectrometry-based techniques are the most commonly used platform in untargeted metabolomics studies due to high sensitivity, high selectivity, and broad dynamic range. Chromatographic separation prior to mass detection enhances data quality (by reducing matrix effects and ionization suppression), increases metabolome coverage and provides additional information for metabolite identification. Pooled samples are usually run in between samples as a quality control check.

iii. Data processing: Due to the global scope of its analysis, untargeted metabolomics produces an extremely complex data set that makes the manual examination of the metabolite features impractical. The data processing in this type of metabolomics approach, therefore, requires the assistance of bioinformatic tools. These metabolomics software typically provide algorithms for retention time alignment between multiple analyses, data filtering, metabolite feature detection (mass to charge value and chromatographic retention time), and relative quantification (chromatographic peak area).

iv. Data analysis: Quality assurance is the first important step in data analysis. Redundant metabolite features (e.g. isotopes and adducts) and unreliable features (poor analytical reproducibility in the quality control samples or elute in ion suppression regions) should be removed. Analysis of quality control samples should also be performed to ensure good instrumental reproducibility throughout the period of analysis. Finally, multivariate (supervised or non-supervised) and univariate (e.g. t-test, fold change) analyses are used to determine the
metabolite features that are altered significantly in the treatment groups versus the control.

v. Metabolite Identification: The metabolites of interest are identified based on authentic standards, or their identity can be tentatively assigned based on matching their accurate mass/empirical formula within metabolite databases (putative identification), such as METLIN [48], Human Metabolome Database [49] and LIPID MAPS [50]. In all cases, metabolites should be annotated based on their characteristic mass-to-charge value and retention time. Despite recent advances in metabolomics technologies, metabolite identification is still the bottleneck step in untargeted metabolomics. Untargeted metabolomics usually produces a large number of unknown metabolites, many of which are putatively identified due to expensive or unavailable authentic standards. Mass accuracy is therefore very important in de novo metabolite identification. As such, in addition to instrumental mass calibration, acquired mass spectra should also be calibrated internally. Internal standards can be added to samples or endogenous compounds such as sodium formate adducts (endogenous sodium is from cells or bio-fluid extracts and formic acid is from the mobile phase) can be used as internal calibrants. The internal calibration can significantly enhance mass accuracy for de novo metabolite identification in metabolomics research.

vi. Data interpretation: Metabolomics data interpretation is the task of understanding the biological functions and system-level effects of the metabolites of interest. Functional analysis and pathway-mapping tools are usually used to determine the interconnectivity of metabolites within metabolic pathways in relation to an aberrant process or a pathological phenotype. These analyses provide insights into the molecular mechanisms underlying a pathological or aberrant condition. Further mechanistic studies could also be done in vitro or in vivo to further explore the biological effects of the metabolites of interest.

8. CASE STUDIES

Metabolomics technologies have been applied to cardiovascular research over the past decade [51, 52]. However, the use of the untargeted approach is still relatively new in this field [53]. Here, we describe five case studies that employ untargeted metabolomics approaches to identify biomarkers for cardiovascular risk [54], myocardial ischemia [55], transient ischemic attack [56], incident coronary heart disease (CHD) [57], and MI risk prediction [58]. These are just a few examples in the literature that cover a broad range of topics in the cardiovascular area.

8.1. Cardiovascular Risks and Atherosclerosis

Hazen and colleagues used untargeted metabolomics to identify plasma metabolites that predict increased risk for CVD [54]. They investigated the plasma metabolomic profiles of 50 stable patients undergoing elective coronary angiography who subsequently had a major adverse cardiac event including nonfatal acute MI, cerebrovascular accident or death. The plasma profiles of these patients were compared to age- and gender-matched controls. Using a LC-MS system, 40 metabolites out of over 2000 monitored metabolites were found to be significantly different between the two groups (using Bonferroni correction and trend criteria). These 40 metabolites were validated in an independent cohort and 18 of them showed consistent and significant associations with the development of major adverse cardiac events. Further structural identification using NMR and tandem MS technologies identified three metabolites, choline, trimethylamine N-oxide, and betaine, which correlated strongly with one another, suggesting a potential relationship via a common metabolic pathway. These three metabolites are involved in the choline and phosphatidylcholine metabolism.

In the follow-up mechanistic studies, mice supplemented with dietary choline showed accelerated development of atherosclerosis. Furthermore, suppression of intestinal microflora inhibits this accelerated development. This is one of the very first studies that identified a direct link between the intestinal microflora, dietary phosphatidylcholine, and CVD risk. The results of this study suggest that targeting the pathogenic metabolic pathway (i.e. choline metabolism) and/or modulating the microbial composition may be a novel therapeutic approach for the prevention and treatment of atherosclerotic CVD. This well-designed translational study demonstrates the power of untargeted metabolomics as a discovery tool for novel biomarkers and pathogenic metabolic pathways.

8.2. Transient Ischemic Attack

Stroke is the leading cause of acquired neurologic incapacity. In approximately 20% cases, stroke is preceded by a transient ischemic attack, which provides an excellent opportunity for intervention and stroke prevention. Several biomarkers have been proposed for the diagnosis of transient ischemic attack including C-reactive protein and copeptin. However, the validity of these proposed biomarkers has been debated [59]. Purroy and colleagues employed untargeted metabolomics to identify novel candidate biomarkers for stroke recurrence (SR) [56]. An LC-MS system was used to analyze plasma samples from transient ischemic attack patients (n=131) recruited within 24 hours after the onset of symptoms. A total of 35 patients had an SR.

Pattern analysis and metabolomic profiling revealed specific biomarkers of SR and also large-arterial atherosclerosis. They found 94 metabolite features that are significantly different between SR and non-SR patients. Among these metabolites, several species of lysophosphatidylcholine (LysoPC) including LysoPC(16:0) and LysoPC(20:4) were identified to be significantly associated with SR. LysoPC (22:6) was also determined to be a potential biomarker of large-arterial atherosclerosis. Metabolomics pattern recognition using multivariate statistics differentiated individuals who present early (less than three months) recurrence, compared
to non-SR and late SR subjects. These results were further validated in an independent cohort (n=162, with 37 cases of SR) to ensure robustness of the resulting candidate biomarker. This study demonstrated the feasibility and utility of metabolomics approach to reveal novel biomarkers for SA after a transient ischemic attack.

8.3. Incident Coronary Heart Disease

Comprehensive analyses of circulating metabolites in large prospective cohorts can lead to improved prediction and better pathological understanding of CHD. Ingelsson and colleagues performed untargeted metabolomics profiling in CHD-free individuals (n=1,208 with 131 CHD events in 10-year median follow-up) for baseline measurement [57]. The defined CHD incidents are acute MI and unstable angina. The aims of this study were to identify novel biomarkers of CHD and delineate the potential causal effects of those metabolites showing a strong association.

Out of the 10,162 LC-MS features detected, the authors identified 32 unique metabolites associated with CHD incidence after false discovery rate correction. These metabolites were further validated in an independent cohort (n=1,670 with 282 CHD events in 3.9-year median follow-up), of which only four metabolites were found to be significantly associated with CVD risk. LysoPC (18:1), LysoPC(18:2), and sphingomyelin (28:1) were found to be associated with lower risk, whereas monoglyceride (18:2) was associated with higher risk. LysoPCs were also found to be negatively associated with body mass index and C-reactive protein. Using Mendelian randomization analysis, they also identified a causal effect for monoglyceride (18:2). This study is one of the largest untargeted metabolomics studies in relation to incident CHD.

8.4. Myocardial Ischemia

Timely and accurately recognition of myocardial ischemia is critical for both diagnosing coronary heart disease and selecting appropriate therapeutic interventions. The identification of metabolic alterations that are associated with myocardial ischemia can lead to the discovery of clinical biomarkers for diagnostic outcomes. Gerszten and colleagues employed untargeted metabolomics to identify and characterize novel biomarkers and pathways associated with myocardial ischemia [55]. Blood samples were collected from patients before and after exercise stress with myocardial perfusion imaging. The plasma metabolomics profile of patients with evidence of inducible ischemia (n=18) was compared to the control without evidence of ischemia. Myocardial ischemia was determined using single-photon emission computed tomography myocardial perfusion imaging.

Metabolite levels were analyzed before and after stress testing, using high-performance LC-MS system, to identify potential novel biomarkers of coronary ischemia. Six metabolites including γ-aminobutyric acid, uric acid, citric acid, and several unidentified metabolites were found to be associated with inducible ischemia. Furthermore, inducible ischemia patients could be effectively differentiated from controls by monitoring changes in these six metabolites. Subsequently, functional pathway analysis of identifiable metabolites revealed that members of the citric acid pathway were significantly overrepresented. Such metabolites and their pathways may ultimately serve as targets for therapeutic intervention or as biomarkers for diagnostic methods.

8.5. MI Risk Prediction

Identifying individuals at an increased risk for MI provides a great opportunity for timely prevention and/or treatment. High sensitivity C-reactive protein is a promising biomarker for MI, however, it shows modest added value to conventional risk scores [60, 61]. Also, genetic risk scores have not substantially improved the predictive value of established MI [62]. Therefore, there remains a need for predictive biomarkers of MI. In a study performed by Peters and colleagues, three independent prospective cohorts (one discovery cohort and two validation cohorts) were used to identify biomarkers for MI and evaluate their involvement in the pathogenesis of MI [58].

In the discovery cohort (n=1,342 with 67 MI), 16 metabolites were found to be associated (p<0.05) with incident MI. These metabolites, which belong to the classes of amino acids, LysoPCs and PCs, showed high intraclass correlations and low interclass correlations. Among the 16 metabolites, arginine, LysoPC(17:0) and LysoPC(18:2) were found to be significantly (p<0.0003) associated with MI in the two validation cohorts (n=661 with 112 MI and n=254 with 87 MI). Increased level of arginine and decreased levels of LysoPC(17:0) and LysoPC(18:2) were associated with increased risk of incident MI. In combination, these three metabolites increased the predictive value of the Framingham risk score in predicting CVD risks.

Arginine, LysoPC(17:0) and LysoPC(18:2) were also found to strongly associated with high sensitivity C-reactive protein, which is a measure of systemic inflammation. These three metabolites attenuated the association between high sensitivity C-reactive protein and MI, suggesting a potential link between these metabolites and systemic inflammation. These inflammation-associated metabolites could be new targets for the investigation of the inflammation-driven pathophysiology of MI.

9. SUMMARY AND FUTURE PERSPECTIVES

The metabolome is an extremely complex mixture composed of many classes of metabolites with differences in bioavailability and chemical diversity. Over the past decade, the field of metabolomics has expanded rapidly, thanks to the advances in analytical techniques and biochemoinformatic tools. Nevertheless, there is always a demand for a better and more sophisticated analytical platform that is capable of detecting more diverse metabolites with greater dynamic range. The rapid development of metabolite databases such as Human Metabolome Database with over 40,000 metabolites has greatly assisted in the identification of unknown metabolites and understanding of the biochemical
and clinical information of the metabolites [63]. However, there still remain a large number of ‘unknown unknown’ metabolites, which could potentially provide great insights into novel metabolic alterations and mechanisms of a disease. Therefore, there remains a need for future advancement in metabolomic techniques including analytical platforms, bioinformatic tools, and metabolite databases.

Biomarkers are capable of indicating a variety of health or disease characteristics, including disease traits, disease state, or disease rate [64]. Defining a normal and/or an abnormal range of biomarker values is a critical step in establishing a clinical biomarker. It is important to characterize the distribution of the markers in multiple communities and laboratories, in order to take into account the variability in inter- and intra-individual, inter- and intra-laboratory, and sample processing. Two approaches are typically used to define abnormal biomarker levels, including reference limit and discrimination limit [65]. Reference limits are statistically derived cut-points (typically 95th or 97.5th percentile), based on the distribution of values in reference samples such as healthy individuals that are free of the disease of interest. A discrimination limit is used to separate patients with and without the disease by evaluating the degree of overlap between the profiles of disease patients and healthy controls. Abnormal cut-points of biomarkers can be varied depending on many factors including the consequences of missing disease (false negative) versus misclassifying non-diseased individuals (false positive). These cut-points need to be determined and validated prior to moving to next phases of biomarker development.

Biomarker discovery in humans is extremely complicated due to the genetic diversity in human population and multiple confounding factors including age, diet, medications, and comorbidities. As such, a combination of learning/discovery and validation cohorts is strongly recommended for metabolomics-based discovery studies. In addition, the discovery of biomarkers in human often requires the ability to differentiate subtle clinical symptoms or phenotypes, not only between patients versus healthy individuals, but also within the patient group. This impedes the use of large cohort studies in the identification of specific diagnostic biomarkers of a disease such as atherosclerosis, ischemic and hemorrhagic stroke.

**CONCLUSION**

Untargeted metabolomics can simultaneously uncover many key metabolites and important biological pathways that are associated with a pathological condition. Therefore, metabolomics profiling allows for the classification of individuals based on their molecular phenotype or “metabotype”, not their visible phenotype. The identification of specific blood-borne metabolites associated with different types of CVD and cardiovascular risks could lead to the establishment of a relatively non-invasive, effective and cost-efficient platform to diagnosis CVDs (and other diseases). These metabolomics-based biomarkers can also be used to identify those at risk for major adverse cardiac events, independent of traditional risk factors. In conclusion, metabolomics-based discovery has shown great potential in cardiovascular research including assessing disease risk, diagnosing the disease, defining pathological mechanisms and identifying therapeutic targets.

**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD</td>
<td>Coronary Heart Disease</td>
</tr>
<tr>
<td>CVD</td>
<td>Cardiovascular Disease</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid Chromatography</td>
</tr>
<tr>
<td>LDL</td>
<td>Low-Density Lipoprotein</td>
</tr>
<tr>
<td>MI</td>
<td>Myocardial Infarction</td>
</tr>
<tr>
<td>MS</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>PC</td>
<td>Phosphatidylcholine</td>
</tr>
<tr>
<td>SR</td>
<td>Stroke Recurrence</td>
</tr>
</tbody>
</table>

**CONSENT FOR PUBLICATION**

Not applicable.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest, financial or otherwise.

**ACKNOWLEDGEMENTS**

This research was supported by operating grants from the Canadian Institutes of Health Research (MOP62910 and MOP142248) and the Heart and Stroke Foundation of Canada (G-17-0017029). V.T.D. is supported by an International Ontario Graduate Scholarship.

**REFERENCES**


