HHS Public Access

Author manuscript

Nat Metab. Author manuscript; available in PMC 2024 June 10.

Published in final edited form as:

Nat Metab. 2023 October ; 5(10): 1656–1672. doi:10.1038/s42255-023-00903-x.

Metabolomic Epidemiology Offers Insights into Disease Aetiology

Harriett Fuller[1], **Yiwen Zhu**[2], **Jayna Nicholas**[3], **Haley A. Chatelaine**[4], **Emily M. Drzymalla**[5], **Afrand K. Sarvestani**[6], **Sachelly Julián-Serrano**[7], **Usman A. Tahir**[8], **Nasa Sinnott-Armstrong**[1], **Laura M. Raffield**[3], **Ali Rahnavard**[6], **Xinwei Hua**[9], **Katherine H. Shutta**[10,11], **Burcu F. Darst**[1,*]

[1]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA

[2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

[3]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC

[4]National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD

[5]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC

[6]Computational Biology Institute, Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, The George Washington University, Washington, DC

[7]Department of Public Health, University of Massachusetts Lowell, Lowell, MA

[8]Department of Cardiology, Beth Israel Deaconess Medical Center, Boston, MA

[9]Department of Cardiology, Peking University Third Hospital, Beijing, China

[10]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[11]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

Abstract

Metabolomic epidemiology is the high-throughput study of the relationship between metabolites and health-related traits. This emerging and rapidly growing field has improved our understanding of disease aetiology and contributed to advances in precision medicine. As the field continues to develop, metabolomic epidemiology could lead to discoveries of diagnostic biomarkers predictive of disease risk, aiding in earlier disease detection and better prognosis. In this review, we discuss key advances facilitated by the field of metabolomic epidemiology for a range of conditions, including cardiometabolic diseases, cancer, Alzheimer's disease, and COVID-19, with

*Corresponding Author Burcu F. Darst, 1100 Fairview Ave N, Seattle, WA, 98109 USA, bdarst@fredhutch.org.

Author contributions

All authors (HF, YZ, JN, HAC, EMD, AKS, SJ, UAT, NS, LMR, AR, XH, KHS, BFD) conducted literature searches for article content, contributed substantially to discussions around article content, and drafted, reviewed, and revised the manuscript. HF, HAC, and AR led the creation of manuscript figures with input and assistance from all other authors.

Competing interests

The authors declare no competing financial interests.

a focus on potential clinical utility. Core principles in metabolomic epidemiology, including study design, causal inference methods, and multi-omic integration, are briefly discussed. Future directions required for clinical translation of metabolomic epidemiology findings are summarised, emphasising public health implications. Further work is needed to establish which metabolites reproducibly improve clinical risk prediction in diverse populations and are causally related to disease progression.

1. Introduction

Metabolomic epidemiology is the study of the relationship between a high-throughput set of small-molecules collectively known as the human metabolome and health-related traits in population-based epidemiologic studies. The “metabolome”, coined in 1998, is the complete set of metabolites synthesised by a biological system[1] (see Box 1 for common metabolomic epidemiology terminologies). Although “metabolomic epidemiology” was broadly introduced in 2021[2], such investigations have been ongoing for over two decades, as rapid technological developments have made the measurement of small molecules more efficient, accurate, and accessible for clinical and research applications.

Findings from this emerging field have already offered timely insights into disease aetiology, early detection, and progression, which could inform preventive, screening, and treatment strategies (Fig. 1). As metabolites represent the end product of many biological processes and are sensitive to environmental exposures, changes in metabolite levels could indicate disease risk at early subclinical stages of disease when prevention is still possible. For instance, a recent study reported that metabolomic states, or profiles, derived from 160 circulating metabolites were associated with incidence rates of diseases including coronary heart disease, type 2 diabetes (T2D), dementia, chronic obstructive pulmonary disease, liver disease, and lung cancer, with metabolomic states significantly improving the discriminative performance of established clinical predictors[3].

Several notable recent efforts have advanced the field of metabolomic epidemiology. The COnsortium of METabolomics Studies (COMETS) was established in 2014 to promote large-scale human metabolome collaborations[4], currently including >380,000 participants from 79 cohorts. COMETS has informed the harmonization of metabolites between platforms and is anticipated to further our understanding of how metabolites relate to disease aetiology and progression. The Finnish Institute for Health and Welfare (THL) Biobank has metabolomic data for >40K participants, with early investigations identifying, for example, metabolic differences between risk of peripheral artery disease versus coronary artery disease (CAD)[5,6]. In 2023, the UK Biobank publicly released metabolomic data on 280K participants, with medical records available for >700 common diseases, with previous data releases providing disease risk insights beyond commonly investigated cardiometabolic conditions[7]. Further, the NHLBI Trans-Omics for Precision Medicine (TOPMed) programme is generating metabolomic data on >37K diverse participants, with plans for longitudinal data generation. Lastly, although standard quality control guidelines for untargeted metabolomics are currently lacking, the metabolomic Quality Assurance and Quality Control Consortium recently provided guidance on the reporting of quality control

and assurance procedures in untargeted metabolomic studies to increase transparency and reproducibility[8].

In this review, we examine the biological insights provided by the growing field of metabolomic epidemiology with regards to prevention, screening, diagnostics, and treatment of several health-related traits that have been commonly examined in the field (Fig. 2). Study design considerations are briefly presented to provide context for interpreting findings. We discuss potential clinical applications for metabolomic epidemiology to advance disease treatment and healthcare in the coming years.

2. Study Design Considerations

Study design directly impacts research scope and interpretation of findings (Fig. 1). This is particularly important in metabolomic epidemiology due to the sensitivity of metabolites to disease onset, confounders or effect mediators (e.g., age, population, body mass index (BMI), fasting status), and technical factors related to sample collection and processing (e.g., sample handling, batch, stochastic drift). Among the most common study designs are case-control studies. Although powerful and cost-effective, case-control studies are particularly prone to selection bias, which can occur when the control group is not representative of the source population. Matched case-control studies are used to control for potential confounders prior to the collection of samples. However, such study designs limit the ability to directly investigate the effect of the matching variables (e.g., age, sex, population) on the outcome, as matching distorts this relationship[9], although matching factors can be examined as effect modifiers. As such, matching factors should not include secondary exposures of interest.

While cross-sectional studies are limited to evaluating correlative associations, as biospecimen collection and outcome ascertainment occur concurrently at a single time point, prospective cohorts with pre-diagnostic metabolomic measurements could facilitate time-to-event investigations. Longitudinal repeated sampling enables investigations of metabolite trajectories that may be relevant to disease risk or other unrelated sources of variation, ultimately providing more stable association estimates. Generating metabolomic data in sampled longitudinal cohorts can be cost-prohibitive; however, nested case-control or case-cohort studies that select cases and controls within a cohort provide an efficient way to evaluate pre-diagnostic metabolites in the pathogenesis of disease. Although randomised controlled trials are often based on smaller sample sizes and cost-prohibitive, they provide rigorous causal evidence, for instance, between an intervention and the metabolome. A detailed review of study designs is beyond the scope of this article; previous publications provide further insights[10].

Another important study design consideration includes metabolomic technology selections, such as nuclear magnetic resonance (NMR) and mass spectrometry (MS), as metabolite measurements vary by methodology and platform. These include targeted approaches that profile select metabolites and untargeted approaches that profile all measurable metabolites within the range of the specific platform. While the identity of a subset of metabolite peaks

may be known, most peaks are unknown, necessitating significant downstream efforts in metabolite peak identification.

Platform considerations include metabolite separation techniques and determining the class and/or polarity of metabolites detected in the study. Technology selections also impact the interpretation of metabolite levels. Targeted methods can be quantitative, providing absolute concentrations using stable isotope labeled standards and calibration curves. Although these methods provide precise quantification, authentic standards are not available for all metabolites, thus limiting metabolite coverage. Further, quantitative methods should also consider protocols specific to the tissue type being analysed. Untargeted methods provide broader coverage, but are semiquantitative, reporting metabolite peak intensities, and measurements are highly dependent on pre-analytic methods. In-depth reviews of these technologies can be found elsewhere[11]. While this review focuses predominantly on circulating metabolites, biospecimen tissue type also influences the realm of research questions and interpretations of study findings, given that tissues of interest vary by disease.

Given the variety of measurement techniques, factors that impact metabolite measurements, quality control procedures, and study designs, replication of metabolomic epidemiology findings has been particularly challenging. As such, strategies to replicate findings, assess generalizability to different populations, and triangulate evidence (see “5.4. Triangulation and validation”) should be carefully considered at the study design stage of an investigation.

3. Metabolomic Epidemiology and Disease Aetiology

Here we discuss major metabolomic epidemiology findings and their clinical and translational implications, particularly highlighting untargeted investigations. Sections 3.1–3.7 focus on traits that have been more commonly investigated in metabolomic epidemiology (Fig. 2) and/or yielded consistent findings across studies to date (Fig. 3), while emerging traits that have not been studied as extensively are briefly discussed in sections 3.8–3.11.

3.1. Adiposity

Heightened adiposity increases risk of numerous health conditions, including T2D, cardiovascular disease (CVD), and numerous cancers. Despite this, mechanisms linking adiposity to adverse health outcomes are not fully understood, although the metabolome likely plays a key role[12]. Metabolomics has been used to predict adiposity and BMI[13], improve our understanding of metabolic dysregulation resulting from increased adiposity[14], and predict the success of obesity treatment[15].

Metabolic perturbations resulting from increased adiposity are systemic and wide-reaching, including changes in glucose, lipid metabolism, and low-grade chronic inflammation, with strong associations consistently observed for branched chain amino acids (BCAAs)[15]. An untargeted study found that nearly a third of the metabolome was associated with BMI, particularly lipids, amino acids, and peptides[12]. Many of these metabolites were also associated with increased insulin resistance (including tyrosine, alanine, kynurenate, gamma-glutamyltyrosine, phospholipids, and glucose) and as such, could

mediate the association between adiposity and metabolic health conditions[12]. However, BMI insufficiently predicts adiposity and its health complications, and metabolomics provides an opportunity to improve such predictions. For example, comparing BMI matched European ancestry individuals, an obese versus healthy metabolome defined by 49 serum metabolites was associated with 2- to 5-fold increase in cardiovascular events[12].

One key determinant of negative health outcomes resulting from increased adiposity is the distribution of adipose tissue, a metabolically active organ that has a substantial impact on the metabolome. When subcutaneous adipose tissue fails to sufficiently expand with dietary triglyceride consumption, visceral and ectopic fat depositions form, leading to inflammatory dysregulation and increased insulin resistance[16]. Understanding fat distribution hence plays a key role in understanding adiposity-related health outcomes; however, clinically implementing anthropometric measures remains cost prohibitive. Establishing metabolic profiles of adiposity may better inform clinical management of adiposity. Metabolomic epidemiology studies have found that plasma steroid sulfates and amino acids were characteristic of visceral and subcutaneous adiposity in individuals living with obesity without insulin resistance[17]. Future prospective studies may investigate whether these metabolites are protective against insulin resistance or markers of prodromal symptoms and provide intervention targets.

Individuals living with obesity can be characterised as metabolically unhealthy overweight and obese (MUHO) or metabolically healthy overweight and obese (MHO). MHO is typically defined as obesity with the absence of metabolic syndrome and insulin resistance[15] and may make up 20–40% of adults living with obesity, with ~50% of those with MHO expected to progress to MUHO[12]. Future large-scale metabolomic epidemiology studies may illuminate metabolic pathways characterising MUHO, leading to more precise MUHO and MHO definitions and improved obesity treatment.

3.2. Cardiovascular disease

CVD is highly heterogeneous, encompassing a range of conditions linked to atherogenic, proinflammatory, and thrombotic mechanisms. Metabolic dysregulation in CVD has been observed since the early 1900s[18], when cholesterol was noted as a putative driver of atherosclerosis, spurring decades of research that established cholesterol as a cause of CVD[19]. Therapies targeting cholesterol metabolism are first-line treatments for atherosclerosis and have greatly reduced atherosclerotic CVD mortality[20], demonstrating the value of metabolic approaches in studying and treating complex diseases.

Early efforts to characterise CVD-associated metabolomic alterations reported increases in inflammatory lipids, including carnitines, phosphatidylcholines (PCs), and fatty acids, and amino acids[21,22]. While these initial studies were primarily based on semi-targeted approaches in small prospective cohorts or case-control studies, many associations proved robust in expanded untargeted investigations across thousands of individuals. For example, an early untargeted metabolomic study with a limited sample size found that trimethylamine N-oxide (TMAO), a downstream microbial metabolite produced from dietary choline and carnitine, was positively associated with CVD risk[22]. This was validated in larger studies, including an untargeted meta-analysis of venous thromboembolism[23] and a meta-analysis

estimating that higher circulating TMAO levels were associated with 23% increased CVD risk[24]. Although it remains up for debate whether TMAO is causally associated with CVD[25], functional studies provide *in vivo* evidence that TMAO may accelerate atherosclerosis progression via suppression of reverse cholesterol transport in macrophages[26]. Similarly, initial positive associations between BCAAs and CVD were confirmed in multiple larger metabolomic studies[23,27], with preliminary evidence for a causal relationship between BCAAs and CAD[28].

Large untargeted metabolomic studies continue to identify additional metabolite-CVD associations, including nucleoside metabolites and steroid hormones[29–31]. Additional efforts to profile metabolite alterations associated with specific CVD outcomes will be critical for identifying disease subtypes within broad CVD event categories. For instance, a prospective study of ischemic stroke identified two long chain fatty acids (tetradecandioate and hexadecandioate) that were specifically associated with the cardioembolic stroke subtype[32]. Additionally, mediation analyses may help disentangle metabolite-related aetiologic mechanisms contributing to disease. For instance, plasma levels of organic acid dimethylguanidino valerate (DMGV) were positively associated with both incident CAD and T2D and inversely correlated with healthy dietary factors and exercise, the latter of which was supported by a subsequent *in vivo* study[33,34]. Subsequently, another study established that DMGV was positively associated with incident ischemic stroke, which was partially mediated by diabetes mellitus (13.0%) and hypertension (13.2%)[31]. Collectively, these studies suggest that associations between DMGV and CVD outcomes may capture a shared disease pathway between CVD and diabetes, influenced by lifestyle factors. Further investigation is needed to determine whether lifestyle factors mediate DMGV-associated risk of CVD and T2D.

CVD has many shared metabolic pathways (notably, organic acids), both within CVD-related and non-CVD traits, particularly adiposity (Fig. 3). However, the lack of shared pathways between non-CVD traits likely reflects the fewer number of studies conducted, and consequently, fewer significant findings for other traits. As the field rapidly grows, shared mechanisms between traits will likely become more apparent.

3.3. Type 2 diabetes

T2D is an increasingly prevalent metabolic condition associated with complications including retinopathy, cardiovascular disease, and kidney disease[35]. One of the most prominent metabolite classes associated with T2D is BCAAs, with positive associations consistently identified across large metabolomic studies[36,37]. T2D risk prediction models have been improved by incorporating metabolites with traditional T2D risk factors, such as sex, age, parental history of T2D, fasting glucose, BMI, high-density lipoprotein, triacylglycerols (TAGs), and blood pressure[38,39]. For example, adding 19 metabolites enriched for nitrogen metabolism to traditional T2D risk factors significantly improved discriminative ability of T2D risk, with[38] higher genetically predicted glycine levels associated with an 11% reduction in T2D risk and lower genetically predicted phenylalanine levels associated with a 60% increase in T2D risk[38]. These findings suggest potentially

causal relationships between the nitrogen metabolism pathway and T2D, providing mechanistic insights into disease onset.

Metabolites have also been used to investigate progression from prediabetes (i.e., the state above normal glucose tolerance but below the T2D threshold) to T2D and T2D-related complications. For instance, alanine, glutamate, and palmitic acids were higher in individuals with prediabetes than those with T2D[40]. Further, the addition of 13 metabolites to traditional T2D risk factors improved prediction of progression from prediabetes to T2D[41]. Serum levels of cyclohexylamine, 1,2-distearoyl-glycero-3-phosphocholine, piperidine, N-acetylneuraminic acid, and stearoylethanolamine have been positively associated T2D complications, including retinopathy and kidney disease[42]. Additional research is needed to replicate these findings and determine their clinical significance.

3.4. Disorders of inborn errors of metabolism

Numerous rare metabolic diseases are defined by their lack, or modification, of critical metabolic enzymes and regulatory systems. These inborn errors of metabolism (IEMs) have long been studied in the context of metabolomic data, with early metabolomic genome-wide association studies (GWAS) confirming a strong relationship between known IEM genes and their corresponding and derivative metabolites[43–46]. Further, extensive work has characterised the metabolic effects of familial hypercholesterolemia variants, a particular class of IEMs[47,48]. As metabolomic technologies have matured, researchers have discovered IEMs using metabolomic outliers, providing an opportunity to connect IEMs to pharmacological interventions[49]. The advent of large cohorts with metabolomic and genetic data has dramatically increased the scale of known IEM-associated metabolites. As many of these still require additional interpretation for clinical relevance, researchers have begun synthesizing findings across studies using databases[50]. Consequently, efforts are underway to clinically apply these discoveries[51,52].

3.5. Infectious disease: COVID-19

Coronavirus Disease 2019 (COVID-19), caused by SARS-CoV-2, has constituted a global pandemic for over three years. The devastating public health effects of COVID-19 have necessitated extensive research to mitigate and prevent infection. Untargeted plasma metabolomics has been integral in developing hypotheses about the metabolic underpinnings of COVID-19 viral pathophysiology and identifying infection and severity biomarkers[53].

Amino acids citrulline, histidine, proline, and tryptophan have been consistently negatively associated with COVID-19 severity[53–59]. Meanwhile metabolites from the kynurenine pathway of tryptophan metabolism have been positively associated with COVID-19 severity[53–57], potentially reflecting increased inflammation in COVID-19. Creatinine has also been positively associated with COVID-19 severity[53,54,57], potentially reflecting metabolic pathways related to renal dysfunction, a possible therapeutic target for severe COVID-19. Cytosine and uridine metabolites have been reported to have opposite associations with COVID-19 severity, with cytosine positively and uridine negatively associated with disease severity[53,54,56]. This has been suggested to reflect viral replication,

as SARS-CoV-2 has low levels of cytosine, which may play a key role in SARS-CoV-2 pathology[60]. Thus, cytosine and uridine metabolism may pose a targetable pathway for decreasing COVID-19 severity.

Bile acids[58] and PCs/sphingomyelins (SMs)[59] have been positively and negatively associated with COVID-19 severity, respectively. Elevated bile acids may reflect metabolic and/or liver dysfunction that predisposes individuals to severe COVID-19[53,58]. However, decreased PCs and SMs may be related to pathophysiology, as both are components of cell membranes upon which ceramide rafts are required for SARS-CoV-2 to bind to ACE2 receptors[61]. Thus, decreasing conversion of PCs and SMs to ceramides for ACE2 presentation may be a targetable pathway for decreasing SARS-CoV-2 infection.

Untargeted metabolomic studies have provided valuable evidence for targetable pathways and potential biomarkers related to COVID-19 severity and serve as a model for future infectious disease outbreaks. Additional untargeted studies are warranted and ongoing to identify additional COVID-19 biomarkers and reduce the burden of viral pandemics on healthcare systems.

3.6. Cancer

Metabolic alterations are often hallmarks of cancer. The Warburg effect is a well-known alteration in glucose metabolism, in which tumours catabolize glucose via glycolysis rather than the tricarboxylic acid (TCA) cycle even in the presence of oxygen. This produces increased lactate levels[62], a metabolic marker of cancer in both tumour and circulating tissue[63], that serves as an intermediate fuel source in the TCA cycle[64,65]. Emerging work indicates that alterations in cancer cell metabolism are highly heterogeneous, even among cells cultured under the same nutrient conditions[66,67]. Altered lipid metabolism has also been implicated in cancer aetiology. Hypoxia in the tumour microenvironment increases lipids in the cell and promotes the transition from aerobic to anaerobic metabolism[68]. Metabolomic epidemiology may reflect these processes and offer insights into cancer development and therapeutics.

Untargeted circulating prospective metabolomic studies have identified promising biomarkers of cancer risk. One study found that choline was positively associated with prostate cancer (PCa)-specific mortality[69], consistent with targeted metabolite and dietary studies[70,71]. Choline and the choline-derivative TMAO produced by intestinal bacteria have also been positively associated with risk of overall and aggressive PCa and colorectal cancer (CRC)[72–74], while inverse associations have been observed between choline derivatives PCs and lyso-PCs with risk of overall and aggressive PCa, CRC, breast cancer (BCa), and cancer in general[73–77]. Choline is essential for lipid metabolism, with a proportion of an individual's necessary choline produced by the liver and the remaining obtained from diet, implicating that this is a partially modifiable risk factor. However, investigations in Chinese populations reported that choline intake was inversely associated with risk of CRC, BCa, nasopharyngeal cancer, and cancer in general[78–81]. While the reasons for these discrepancies are uncertain, population differences in dietary choline sources or metabolism may be contributing factors. Such discrepant findings highlight the importance of performing metabolomic epidemiology studies in diverse global populations.

Untargeted circulating metabolomic studies have also reported an inverse association between bile acid tauro-beta-muricholate and PCa risk[73,74]. This was supported by an *in vitro* study reporting that bile acids selectively induce PCa cell death, sparing normal prostate cells[82]. Dysregulation of bile acid metabolism is particularly important in the pathogenesis of hepatocellular carcinoma (HCC). Targeted and untargeted metabolomic studies consistently found that glycine- and taurine conjugated-primary bile acids, including glycocholic acid, taurocholic acid, and glycochenodeoxycholic acid, were positively associated with HCC risk[83–85], while retinol was inversely associated with HCC risk[84,86].

The circulating metabolite perturbations highlighted here, observed up to 20 years before cancer diagnosis or development of lethal disease, improve our understanding of cancer aetiology while demonstrating the potential to improve cancer risk stratification and tools for early detection. These findings are supported by tumoir metabolomic studies, which offer unique and complementary insights into cancer prognosis and treatment. For instance, circulating levels of the amino acid aspartate were positively associated with risk of PCa-specific mortality[87], while prostate tumour aspartate levels were positively associated with risk of biochemical recurrence and *ERG* translocation[88].

Prostate tumour metabolomic profiles reportedly differ by ERG subtypes, with ERG-positive tumours having higher levels of acylcarnitines and metabolites involved in purine metabolism and lower glutathione levels compared to ERG-negative tumours[89]. These metabolites are indicators of oxidative stress, which can lead to DNA damage and plays a major role in PCa development and progression[89]. Pre-diagnostic circulating metabolomic profiles also differ between by ERG or PTEN molecular subtypes, with ERG-positive tumours uniquely enriched for phosphatidylethanolamines, PTEN-loss tumours enriched for amino acids, and PTEN-intact tumours enriched for unsaturated diacylglycerols[90].

Compared to normal tissues, tumours from women with triple negative BCa were enriched for phosphatidylinositols, fatty acids, and ceramides, and metabolomic profiling refined the classification of transcriptomic subtypes (i.e., luminal androgen receptor, basal-like immunosuppressed, immunomodulatory, and mesenchymal-like)[91]. Overall, these metabolomic findings suggest distinct etiologies and presentations of tumour subtypes, which could have important prognostic and treatment implications.

3.7. Alzheimer's disease and related dementias

Although Alzheimer's disease (AD) development is not completely understood, pathological changes that cause AD begin decades prior to its diagnosis[92]. Metabolomics may provide insights into AD aetiology and early risk factors. Higher levels of PCs and SMs have been associated with progression from mild cognitive impairment (MCI) to AD, faster cognitive decline, and changes in ventricular volume[93]. Some of these metabolites, including SM C16:0 and SM (OH) C14:1, have also been positively associated with AD severity[94]. Metabolite panels have also been shown to discriminate between AD and normal cognition. The diagnostic capability of a metabolite panel with six metabolites (arachidonic acid, N,N-dimethylglycine, thymine, glutamine, glutamic acid, and cytidine) was equivalent to diagnoses through clinical interviews[95], which could have important implications for improving diagnostics.

A defining feature of AD is the aggregation of hyperphosphorylated and misfolded tau proteins in neurons, leading to the hallmark neurofibrillary tangles[96]. Cerebrospinal fluid (CSF) metabolomic studies have reported 38 CSF metabolites enriched for pentose and glucuronate interconversions and glycerophospholipids explained ~70% of the variance of total and phosphorylated tau[97,98]. Adding seven of these metabolites to traditional AD risk factors notably improved the predictive ability of AD and MCI[97].

The Alzheimer's Disease Metabolomics Consortium was established to build a large comprehensive metabolomic database and can be queried with AD Atlas, an integrative network-based resource that enables analyses of multi-omic data and AD risk, biomarkers, and endophenotypes[99]. Additional consortium-scale analyses of prospective cohorts will be crucial to establish mid-life metabolomic predictors of late-life dementia.

3.8. Inflammatory bowel disease

Crohn's disease (CD) and ulcerative colitis (UC) are two major forms of inflammatory bowel disease (IBD), a group of chronic, idiopathic gastrointestinal disorders characterised by inflammation of intestinal mucosa. A growing body of evidence is emerging for IBD metabolomics[100]. By profiling stool, plasma/serum, and urine samples, studies have 1) compared metabolomes of CD or UC to healthy controls; 2) developed metabolomics signatures distinguishing between UC and CD; and 3) identified metabolic profiles for disease activity and treatment response.

Circulating metabolomic investigations have consistently demonstrated perturbations of several amino acids. Tryptophan was significantly lower in IBD patients[101] and indicated as a potential biomarker for response to infliximab, a commonly prescribed anti-inflammatory monoclonal antibody medication, in CD patients[101]. Higher isoleucine and lower glutamine, among other amino acid differences, have been observed in patients with CD and UC compared to controls[102,103]. Further, 3-hydroxybutyrate, a downstream metabolite of BCAAs, was upregulated in UC patients compared to controls[104]. Using blood samples collected four or more years prior to IBD onset, bile acids, amino acids, and steroid hormones were associated with CD risk, while fatty acids were associated with UC risk[105]. However, as few studies have characterised the circulating metabolome prior to IBD development, future prospective and longitudinal studies may provide further insights into the pathogenesis of IBD and biomarkers for early detection.

3.9. Chronic kidney disease

Diagnosis of chronic kidney disease (CKD) is based on the estimated glomerular filtration rate (eGFR), using creatinine values or cystatin C, and on albuminuria category[106], which all indicate kidney function. However, these CKD markers are affected by non-renal processes (i.e., nutritional status), suggesting the need for additional diagnosis biomarkers. Although untargeted metabolomic studies of CKD have focused on detecting stage-specific metabolites to improve diagnostic accuracy, a consistent metabolite associated with the five stages of CKD is lacking. Untargeted metabolomic studies have identified amino acids[107–109], nucleotides[107], carbohydrates[107,108], lipids[107], and cofactors and vitamins[107] that were associated with declined kidney function and CKD. CKD progression has been

associated with the amino acid tryptophan[109,110], which was also associated with diabetic glomerulopathy in T2D patients[111]. Since diabetes mellitus is the leading cause for CKD[112], targeting metabolites related to diabetes may also aid in CKD prevention.

Circulating TMAO was positively associated with renal dysfunction in CKD patients in an untargeted metabolomic study[113] and reportedly distinguished late-stage CKD from earlier stages and controls[110]. A targeted metabolomic study found that higher acetylcarnitine levels were associated with 54% lower eGFR and thus increased CKD risk[114]. Higher acylcarnitines have been observed in pediatric patients with CKD[115] and also associated with decreased kidney function[114], and thus, could be promising metabolites for early CKD diagnosis.

3.10. Pregnancy and gestational diabetes

The maternal metabolome associates with a variety of pregnancy related complications. Due to this relationship between maternal and offspring metabolic health, pregnancy provides a unique opportunity to improve the metabolic health of the mother and the neonate. Predictive metabolic profiles for pregnancy related complications could inform preventive interventions for expectant mothers at risk of adverse pregnancy outcomes. Metabolic profiles of gestational diabetes[116], preeclampsia[117] and macrosomia[118] have been characterised and include fatty acids, cholesterols, and triglycerides. A key challenge in studying the maternal metabolome is accounting for the metabolomic perturbations that accompany pregnancy, given the rising energy demands of the foetus[116], necessitating longitudinally measured metabolomics.

3.11. Psychological distress and mental health

An emerging area of research in metabolomic epidemiology is the identification of metabolomic signatures associated with psychological disorders and subclinical levels of distress, including depression, anxiety, or posttraumatic stress disorder (PTSD). Prior studies have suggested associations between lipids and depression[119] and highlighted key pathways implicated in the pathophysiology of mood disorders, such as glutamatergic metabolism and neurotransmission[120]. Literature in PTSD, anxiety, and subclinical distress is relatively sparse, with vast heterogeneity and inconsistency between studies. However, there is suggestive evidence for associations between fatty acids and general distress across disorders[121]. Understanding the metabolomic underpinnings of psychological distress has implications beyond mental health, as these pathways provide a potential mechanism linking chronic distress to heightened risks for cardiometabolic conditions and other ageing-related diseases[122]. Research in diverse population-based samples is needed to identify robust metabolomic signatures of psychological distress.

4. Integrating Metabolomics with Other Omics

In this section, we discuss insights gained by integrating metabolomics with other Omics data, particularly genetics, the microbiome, and the exposome, focusing on key translational insights.

4.1. Metabolomics and Genetics

The heritability of circulating metabolites ranges on average from ~20–50%[43,46,123–126], which is greater than what is often observed for complex traits[127], including whole-blood gene expression, which ranges from ~10–25% on average[128–130]. As such, integrating metabolomics and genomics has strong potential to illuminate metabolic mechanisms and advance precision medicine.

The human metabolome is highly polygenic, with most associated variants leading to minor changes in metabolite levels[43], although some variants can lead to severe IEMs and GWAS signals sometimes cluster near IEM-related genes[131]. Identifying variants that influence metabolites improves our understanding of the biological processes regulating metabolites and are impacted by metabolites, contributing to disease prevention and treatment[43]. Further, genetics provides a notable means of determining the chemical identity or compound class of unknown metabolites, which commonly result from untargeted experiments and pose major challenges to interpreting findings. Multiple metabolites are often associated with the same gene, highlighting shared metabolic and potentially causal pathways[132], which was demonstrated in an investigation where the majority of 336 unknown metabolites associated with genomic loci were able to be successfully annotated using a combination of bioinformatics tools[133].

To date, over 25 metabolomic GWAS have been conducted in European ancestry populations[126,134], resulting in ~ 1,750 independent metabolite-variant associations identified (Fig. 4). Fewer metabolomic GWAS have been conducted in African[133,135], Hispanic[136], Asian[134,137–139], and Middle Eastern[124] populations, typically with smaller sample sizes. Nonetheless, metabolite-variant associations have been identified across the genome in all populations. One of the most notable genomic regions is the fatty acid desaturase (FADS) locus on chromosome 11q12.2, which contains three genes, *FADS1, FADS2*, and *FADS3* (Fig. 4). *FADS1* and *FADS2* encode desaturase enzymes involved in long-chain polyunsaturated fatty acid (PUFA) biosynthesis, while the role of *FADS3* is still unknown. FADS variants are highly associated with PUFAs across tissues and contribute to several diseases, including CAD, cancer, and T2D[140–142]. FADS highlights the importance of diversity in metabolomic GWAS, as this region essentially represents one large linkage disequilibrium (LD) block in individuals of European ancestry, but many smaller blocks in individuals of African ancestry[143], narrowing the range of potentially causal variants to target for therapeutic applications. LD differences in this region likely reflect historically different dietary patterns between populations that acted as selective pressures[143].

Several useful resources exist to explore published GWAS associations with metabolites, including the GWAS catalog[144], mGWAS-Explorer[145], Phenoscanner[146], and PheWeb[147]. The UK Biobank atlas of polygenic risk scores (PRS) summarizes associations between 129 PRS and 249 circulating metabolites[148]. Using this resource, robust associations were identified between an adiposity-related PRS and the majority of metabolites measured, demonstrating its potential to improve our understanding of how genetic risk of complex traits could impact metabolite levels and disease development.

4.2. Host-gut microbiome and metabolome

The gut microbiome plays an important role in disease aetiology, as it reflects lifestyle factors, including diet and exercise, and markers of disease risk, including BMI, insulin resistance, cholesterol, and inflammation and also influences the circulating and peripheral organ metabolome via active and passive nutrient uptake[149]. As metabolomics gained recognition as an important component of the microbiome, the microbiome shifted from being studied from the taxonomic perspective via genomics/metagenomics to integrating metabolomics and metagenomics[150].

Although metabolomics has suggested intriguing findings related to the gut microbiome (including the above-mentioned bile acid and TMAO findings), few published metabolomic epidemiology studies have integrated the gut microbiome. Reasons for this include sample collection challenges, microbiota sample heterogeneity, and the lack of universal approaches to process, analyse, and interpret samples and data. Efforts expected to increase the feasibility of microbiome-metabolomic studies include the Global Natural Product Social Molecular Networking (GNPS) tool[151] that identifies the molecular fingerprint of unknown features in the microbiome and microbial metabolome. For example, GNPS identified a large group of previously unknown bile acids formed from the bacterial conjugation of amino acids to bile acids[152]. Further, collecting samples across the gastrointestinal tract using collection capsules will offer a more comprehensive understanding of the microbiome[153].

4.3. Exposome

The exposome is the complete set of an individual's exposures across the lifespan. Metabolomics is an important tool in studying the exposome, as environmental exposures can be measured via exogenous metabolites (e.g., those influenced predominantly by diet, medications, environment, and lifestyle). Endogenous metabolites (e.g., those influenced predominantly by the genome, epigenome, transcriptome, and proteome) are also thought to be influenced by the exposome, although it is not typically apparent whether a metabolite is endogenous or exogenous in origin. In one application, an environment-wide association study of T2D risk found a protective association for β-carotenes nutrients and negative associations for the phenol lipid γ-tocopherol and the pesticide heptachlor epoxide[154]. Efforts are ongoing to incorporate the exposome into integrated frameworks across molecular omics[155].

5. Advancing Metabolomic Epidemiology Findings

Causal inference is often limited in epidemiologic investigations, making metabolomic epidemiology findings challenging to interpret. In this section, we discuss key examples of moving initial metabolomic findings towards clinical translation, focusing on approaches to investigate whether identified associations are causal or correlative in nature. Sections 5.1–5.4 discuss statistical approaches applied in metabolomic epidemiology that have led to etiologic insights, focusing on the findings from these studies (statistical details provided in Boxes 2–3). Section 5.5 discusses findings from experimental investigations following up on metabolomic epidemiology results.

5.1. Mendelian randomisation

Mendelian randomisation (MR) is a causal inference technique that determines the presence of causal relationships between exposures and outcomes by utilizing genetic variants as a proxy for an exposure (Box 2), making MR most suitable for heritable exposures. Since an individual's genetics are randomized at conception, genetic variants are not subjected to the confounding observed in observational studies and can thus be used as instrumental variables (IVs). MR also bypasses issues of reverse causality, which is particularly important in metabolomics as metabolites are often influenced by the outcome. Large-scale metabolomic MR studies have identified putative causal associations between circulating metabolites and AD[156], autoimmune diseases (type 1 diabetes and inflammatory bowel disease)[157], T2D[158], and cancer (lung, ovarian, breast cancer, and glioma)[159].

MR has several key limitations and assumptions (Box 2), which are particularly impacted by the high correlation between metabolites. As such, MR findings should be interpreted with caution and additional follow up is necessary to validate findings, for example with triangulation of evidence (see "5.4. Triangulation and validation") and by evaluating assumption violations.

5.2. Causal mediation analysis

Causal mediation analysis is a framework for researchers to identify pathways through which metabolites mediate the relationship between exposures and outcomes. These pathways can reveal etiologic insights and intervention targets. High-dimensional methods have been implemented to detect joint mediating effects of multiple metabolites and other omic markers, which led, for instance, to the identification of growth hormone receptor, caffeine metabolism, and valine, leucine, and isoleucine degradation as mediators of the effect of bariatric surgery on glycemia, insulin secretion, and insulin sensitivity, respectively, among T2D patients[160]. The validity and clinical implications of causal mediation analysis should be evaluated based on the plausibility of key assumptions (Box 2).

5.3. Network methods

Metabolites in similar pathways are often correlated, and correlation between metabolites in separate pathways also occurs due to the biochemical principles underlying metabolism[161]. This underlying correlation structure may be leveraged to generate or validate functional hypotheses about metabolic processes by mining network models of metabolomic data.

Correlation networks are a commonly used network modelling approach in metabolomic epidemiology models. Weighted Gene Coexpression Network Analysis (WGCNA) extends the concept of correlation networks by identifying network modules based on weighted edges corresponding to between-metabolite correlations (Box 2). For example, WGCNA was applied to untargeted circulating metabolites to identify six metabolite modules associated with measures of lung function in children with asthma, including one enriched for lipid metabolism[162]. Integrating this module with WGCNA-based gene expression modules led to the identification of an association between asthma and *ORMDL3* and subsequently, a nearby asthma-associated variant, generating a mechanistic hypothesis about the role of genetic variation, gene expression, and lipid metabolism in asthma.

Gaussian graphical models (GGMs, or partial correlation networks) are another type of network model that is commonly applied in metabolomic epidemiology. GGMs have been used to reconstruct metabolic classes[163], and network-based clustering of GGMs was used to propose data-driven gender-specific metabolite modules[164]. Playdon et al. constructed a GGM on 113 diet-related prediagnostic circulating metabolites and identified three metabolite modules associated with ER-positive BCa risk, which mapped to three dietary categories: alcohol, vitamin E, and fats and oils[165].

Although correlation networks and GGMs do not represent causal relationships, they can help develop causal hypotheses for subsequent investigations.

5.4. Triangulation and validation

Given the vast heterogeneity in metabolomic epidemiology regarding measurement techniques, populations, and study designs, it is important to triangulate evidence, meaning that evidence is combined across multiple statistical methods and/or data sources. Integrating results from different study designs, each with distinct sources of bias, enables researchers to assess the validity and generalizability of observed associations, including the strength of causal evidence[166]. For example, one study triangulated evidence from longitudinal observational cohorts and bi-directional MR to elucidate a causal relationship between the essential BCAA leucine and T2D risk[158]. Replication plays a pivotal role in establishing causal evidence and is a crucial validation step, particularly when the discovery sample is relatively small and limited to correlative inference (e.g., cross-sectional studies). When corroborating findings across studies, it is crucial to consider the degree to which differences in findings can be attributed to bias, measurement methods, and populations. To compare and synthesize results qualitatively and quantitatively, performing systematic reviews and meta-analyses may be important (Box 3).

5.5. Linking findings to biochemical and functional implications

Metabolomic epidemiology studies often rely on cross-sectional observational data based on a single tissue, typically plasma or serum. Follow-up experiments in animal models and cell lines are key to investigate biochemical and functional implications of metabolite-disease associations.

For example, an untargeted circulating metabolomic epidemiology investigation discovered and externally validated a positive association between phenylacetylglutamine (PAGln) and risk of major adverse cardiovascular events[167]. Subsequently, the investigators confirmed that PAGln production is dependent on gut microbiota in humans and mice, and a series of follow-up experiments revealed the mechanisms underlying associations between gut-microbial derived PAGln and CVD risk. Among these were *ex vivo* experiments demonstrating that PAGln promoted platelet functions, *in vivo* experiments in arterial injury mice models highlighting that PAGln and phenylacetylglycine (PAGly) can lead to increased rate of thrombus formation, and *in vivo* genetic engineering experiments identifying that gut microbial genes *porA* and *fldH* involved in PAGln production can regulate host platelet function and thrombosis. The authors also found that the use of a β-adrenergic receptor antagonist (β-blocker, propranolol) reduced PAGln-induced

platelet hyper-responsiveness, and carvedilol, a β-blocker used for hypertension and heart failure, reversed the prothrombotic effects of PAGln in mice. Collectively, these follow-up experiments illuminate the pathophysiologic mechanisms underlying the association between PAGln and CVD risk.

In another example, a metabolomic study found that metformin treatment in T2D patients was associated with decreased serum citrulline, which was validated in murine tissues from metformin-treated diabetic mice[168], highlighting mechanistic pathways altered via metformin treatment.

Experimental investigations can provide a mechanistic and complementary understanding of population-based findings and are thus an important aspect of metabolomic epidemiology. Experimental findings can also inform research questions to pursue in population-based studies, reflecting an iterative and interdisciplinary process that can offer dynamic insights into disease aetiology.

6. Future Developments and Clinical Implementation Outlook

Despite being a relatively new field, metabolomic epidemiology has provided critical insights into disease aetiology by linking the metabolome to various chronic diseases and identifying etiologic mechanisms across conditions. However, to fully realise the clinical utility of metabolomics and translate it into effective treatment strategies, several advances and avenues of future research should be pursued.

Currently, sample sizes of untargeted studies range from 100 to ~3,000, and while many findings have been robustly replicated (Fig. 3), many more have not been independently replicated or validated. As technologies improve and larger sample sizes become feasible, power to confidently identify and replicate findings will substantially increase. Concerted efforts to share metabolomic data (e.g., Metabolomics Workbench, dbGaP, BioLINCC, MetaboLights) will also ameliorate these issues. Recent progress in collecting and sharing large-scale metabolomics data in population-based biobank studies, such as the TOPMed, UK Biobank, THL biobank, and China Kadoorie Biobank hold great promise for improving statistical power and reproducibility.

Future research should aim to increase the diversity of participants included in metabolomic studies to ensure that the everyone can benefit from findings. The metabolome has been shown to differ between populations[169,170], and population-specific metabolome associations have been observed, including for atherosclerosis[171] and gestational diabetes[172]. While differences in metabolite-outcome associations may be influenced by exogenous factors that differ between populations (e.g., diet and environmental pollutants), they could also be attributed to genetic ancestry, which often is correlated (though not synonymous) with socially constructed population descriptors. As such, a lack of diversity limits the generalizability of metabolomic epidemiology findings and the potential for clinical translation. Increasing diversity will lead to a better understanding of the role of the metabolome in disease risk and may reduce health disparities.

Another promising avenue for future research is metabolite risk scores (MRS), which represent the cumulative impact of metabolites on a trait. For example, an MRS was found to predict weight gain beyond the predictive ability of clinical covariates or single metabolites alone, indicating a potential clinical application of MRS in risk prediction[173]. Metabolite scores that are characteristic of dietary intake offer a unique opportunity to improve diet assessment and our understanding of how impacts health outcomes. For example, a metabolite score characterising red meat consumption was associated with increased T2D risk[174], while a score characterising adherence to a Mediterranean diet was associated with decreased CVD risk[175].

Metabolomic epidemiology has strong potential to contribute to the development of diagnostic tests, which has traditionally involved a labor-intensive process of selecting a few functionally validated metabolites from a panel of hundreds. Recent studies suggest that metabolomic-based diagnostics could simplify this process, with metabolomic profiles distinguishing between different autoimmune diseases and cancers and predicting disease outcomes[3,176]. Further, a newborn screening study found that untargeted metabolomic profiling led to a six-fold higher diagnostic rate of IEMs compared to traditional screening, which includes a limited number of validated metabolites and metabolic conditions[177]. However, regulatory challenges need to be addressed. A major challenge of clinically translating metabolomic epidemiology findings is the costly and time-intensive process of developing an analytically validated assay, which is conducted in clinical laboratories and typically using targeted platforms with absolute quantification, clinically validating the assay in clinical trials, and clinically implementing the assay, which includes obtaining approvals for medical testing (e.g., from the U.S. Food and Drug Administration)[178].

Beyond challenges with replication, validation, and increasing the diversity of participants included in metabolomic investigations, overcoming several other limitations may further the clinical utility of metabolomic epidemiology findings. These include determining the chemical identity of unknown metabolites resulting from untargeted experiments, improving the harmonization of metabolites across platforms and experiments, reducing costs of untargeted experiments, and comprehensively modelling disease risk in integrative multi-omic investigations.

Over the next decade, rapid technological and computational advances are expected to help address challenges in the field of metabolomic epidemiology and facilitate improved screening, diagnostics, drug development, and disease management. As the field continues to expand, particularly in less frequently studied non-cardiometabolic conditions, and more interdisciplinary collaborations form, expanding the breadth of knowledge that can be gained, a larger proportion of the global population will benefit from the promising field of metabolomic epidemiology.

Acknowledgments

This work was supported by the National Cancer Institute at the National Institutes of Health (grant R00 CA246063 to BFD), the National Science Foundation (grant 2109688 to AR), and an award from the Andy Hill Cancer Research Endowment Distinguished Researchers Program (BFD). KHS was supported by the National Heart Lung and Blood Institute at the National Institutes of Health (grant 2T32HL007427). HAC was supported in part by the

Intramural Research Program of the National Center for Advancing Translational Sciences, National Institutes of Health (ZICTR000410–03).

References

1. Oliver SG, Winson MK, Kell DB & Baganz F Systematic functional analysis of the yeast genome. Trends Biotechnol. 16, 373–378 (1998). [PubMed: 9744112]
2. Lasky-Su J, Kelly RS, Wheelock CE & Broadhurst D A strategy for advancing for population-based scientific discovery using the metabolome: the establishment of the Metabolomics Society Metabolomic Epidemiology Task Group. Metabolomics 17, 45 (2021). [PubMed: 33937923]
3. Buergel T et al. Metabolomic profiles predict individual multidisease outcomes. Nat. Med. 28, 2309–2320 (2022). [PubMed: 36138150]
4. Yu B et al. The Consortium of Metabolomics Studies (COMETS): Metabolomics in 47 Prospective Cohort Studies. Am. J. Epidemiol. 188, 991–1012 (2019). [PubMed: 31155658]
5. Ahola-Olli AV et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia 62, 2298–2309 (2019). [PubMed: 31584131]
6. Tikkanen E et al. Metabolic Biomarker Discovery for Risk of Peripheral Artery Disease Compared With Coronary Artery Disease: Lipoprotein and Metabolite Profiling of 31 657 Individuals From 5 Prospective Cohorts. J. Am. Heart Assoc. 10, e021995 (2021). [PubMed: 34845932]
7. Julkunen H et al. Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. Nat. Commun. 14, 604 (2023). [PubMed: 36737450]
8. Kirwan JA et al. Quality assurance and quality control reporting in untargeted metabolic phenotyping: mQACC recommendations for analytical quality management. Metabolomics 18, 70 (2022). [PubMed: 36029375]
9. Ahlbom A Modern Epidemiology, 4th edition. TL Lash, TJ VanderWeele, S Haneuse, KJ Rothman. Wolters Kluwer, 2021. Eur. J. Epidemiol. 36, 767–768 (2021). [PubMed: 34216355]
10. Chu SH et al. Integration of Metabolomic and Other Omics Data in Population-Based Study Designs: An Epidemiological Perspective. Metabolites 9, (2019).
11. Fiehn O Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. Curr. Protoc. Mol. Biol. 114, 30.4.1–30.4.32 (2016).
12. Cirulli ET et al. Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk. Cell Metab. 29, 488–500.e2 (2019). [PubMed: 30318341]
13. Cirulli ET et al. Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk. Cell Metab. 29, 488–500.e2 (2019). [PubMed: 30318341]
14. Wolfert AI & Oken DE Glomerular hemodynamics in established glycerol-induced acute renal failure in the rat. J. Clin. Invest. 84, 1967–1973 (1989). [PubMed: 2592568]
15. Rangel-Huerta OD, Pastor-Villaescusa B & Gil A Are we close to defining a metabolomic signature of human obesity? A systematic review of metabolomics studies. Metabolomics 15, 93 (2019). [PubMed: 31197497]
16. Neeland IJ, Poirier P & Després J-P Cardiovascular and Metabolic Heterogeneity of Obesity: Clinical Challenges and Implications for Management. Circulation 137, 1391–1406 (2018). [PubMed: 29581366]
17. Koay YC et al. Metabolomics and Lipidomics Signatures of Insulin Resistance and Abdominal Fat Depots in People Living with Obesity. Metabolites 12, (2022).
18. Windaus A Über den Gehalt normaler und atheromatöser Aorten an Cholesterin und Cholesterinestern.: 67, 174–176 (1910).
19. Goldstein JL & Brown MS A century of cholesterol and coronaries: from plaques to genes to statins. Cell 161, 161–172 (2015). [PubMed: 25815993]
20. Centers for Disease Control and Prevention (CDC). Decline in deaths from heart disease and stroke--United States, 1900–1999. MMWR Morb. Mortal. Wkly. Rep. 48, 649–656 (1999). [PubMed: 10488780]
21. Shah SH et al. Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. Am. Heart J. 163, 844–850.e1 (2012). [PubMed: 22607863]

22. Wang Z et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472, 57–63 (2011). [PubMed: 21475195]
23. Jiang X et al. Metabolites Associated With the Risk of Incident Venous Thromboembolism: A Metabolomic Analysis. J. Am. Heart Assoc. 7, e010317 (2018). [PubMed: 30571496]
24. Qi J et al. Circulating trimethylamine N-oxide and the risk of cardiovascular diseases: a systematic review and meta-analysis of 11 prospective cohort studies. J. Cell. Mol. Med. 22, 185–194 (2018). [PubMed: 28782886]
25. Jia J et al. Assessment of Causal Direction Between Gut Microbiota-Dependent Metabolites and Cardiometabolic Health: A Bidirectional Mendelian Randomization Analysis. Diabetes 68, 1747–1755 (2019). [PubMed: 31167879]
26. Koeth RA et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat. Med. 19, 576–585 (2013). [PubMed: 23563705]
27. Holmes MV et al. Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. J. Am. Coll. Cardiol. 71, 620–632 (2018). [PubMed: 29420958]
28. Jiang W et al. Mendelian Randomization Analysis Provides Insights into the Pathogenesis of Serum Levels of Branched-Chain Amino Acids in Cardiovascular Disease. Metabolites 13, (2023).
29. Trainor PJ et al. Systems characterization of differential plasma metabolome perturbations following thrombotic and non-thrombotic myocardial infarction. J. Proteomics 160, 38–46 (2017). [PubMed: 28341595]
30. Lind L, Fall T, Ärnlöv J, Elmståhl S & Sundström J Large-Scale Metabolomics and the Incidence of Cardiovascular Disease. J. Am. Heart Assoc. 12, e026885 (2023). [PubMed: 36645074]
31. Ament Z et al. Nucleosides Associated With Incident Ischemic Stroke in the REGARDS and JHS Cohorts. Neurology 98, e2097–e2107 (2022). [PubMed: 35264422]
32. Sun D et al. A prospective study of serum metabolites and risk of ischemic stroke. Neurology 92, e1890–e1898 (2019). [PubMed: 30867269]
33. Ottosson F et al. Dimethylguanidino Valerate: A Lifestyle-Related Metabolite Associated With Future Coronary Artery Disease and Cardiovascular Mortality. J. Am. Heart Assoc. 8, e012846 (2019). [PubMed: 31533499]
34. Wali JA et al. Nutritional and metabolic regulation of the metabolite dimethylguanidino valeric acid: an early marker of cardiometabolic disease. Am. J. Physiol. Endocrinol. Metab. 319, E509–E518 (2020). [PubMed: 32663097]
35. Zheng Y, Ley SH & Hu FB Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat. Rev. Endocrinol. 14, 88–98 (2018). [PubMed: 29219149]
36. Morze J et al. Metabolomics and Type 2 Diabetes Risk: An Updated Systematic Review and Meta-analysis of Prospective Cohort Studies. Diabetes Care 45, 1013–1024 (2022). [PubMed: 35349649]
37. Wang TJ et al. Metabolite profiles and the risk of developing diabetes. Nat. Med. 17, 448–453 (2011). [PubMed: 21423183]
38. Merino J et al. Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose. Diabetologia 61, 1315–1324 (2018). [PubMed: 29626220]
39. Shi L et al. Plasma metabolites associated with type 2 diabetes in a Swedish population: a case-control study nested in a prospective cohort. Diabetologia 61, 849–861 (2018). [PubMed: 29349498]
40. Long J et al. Metabolite biomarkers of type 2 diabetes mellitus and pre-diabetes: a systematic review and meta-analysis. BMC Endocr. Disord. 20, 174 (2020). [PubMed: 33228610]
41. Ren M et al. Potential Novel Serum Metabolic Markers Associated With Progression of Prediabetes to Overt Diabetes in a Chinese Population. Front. Endocrinol. 12, 745214 (2021).
42. Tomofuji Y et al. Identification of serum metabolome signatures associated with retinal and renal complications of type 2 diabetes. Commun. Med. 3, 5 (2023). [PubMed: 36624208]
43. Shin S-Y et al. An atlas of genetic influences on human blood metabolites. Nat. Genet. 46, 543–550 (2014). [PubMed: 24816252]

44. Suhre K et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature 477, 54–60 (2011). [PubMed: 21886157]
45. Mootha VK & Hirschhorn JN Inborn variation in metabolism. Nature genetics vol. 42 97–98 (2010). [PubMed: 20104246]
46. Long T et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat. Genet. 49, 568–578 (2017). [PubMed: 28263315]
47. Sliz E et al. Metabolomic consequences of genetic inhibition of PCSK9 compared with statin treatment. Circulation 138, 2499–2512 (2018). [PubMed: 30524137]
48. Du Z et al. Low-density lipoprotein receptor genotypes modify the sera metabolome of patients with homozygous familial hypercholesterolemia. iScience 25, 105334 (2022). [PubMed: 36325061]
49. McGregor TL et al. Characterising a healthy adult with a rare HAO1 knockout to support a therapeutic strategy for primary hyperoxaluria. Elife 9, (2020).
50. Mandal R, Chamot D & Wishart DS The role of the Human Metabolome Database in inborn errors of metabolism. J. Inherit. Metab. Dis. 41, 329–336 (2018). [PubMed: 29663269]
51. Steinbusch LKM et al. Targeted urine metabolomics with a graphical reporting tool for rapid diagnosis of inborn errors of metabolism. J. Inherit. Metab. Dis. 44, 1113–1123 (2021). [PubMed: 33843072]
52. Hampe MH, Panaskar SN, Yadav AA & Ingale PW Gas chromatography/mass spectrometry-based urine metabolome study in children for inborn errors of metabolism: An Indian experience. Clin. Biochem. 50, 121–126 (2017). [PubMed: 27784639]
53. Rahnavard A, Mann B, Giri A, Chatterjee R & Crandall KA Metabolite, protein, and tissue dysfunction associated with COVID-19 disease severity. Sci. Rep. 12, 12204 (2022). [PubMed: 35842456]
54. Roberts I et al. Untargeted metabolomics of COVID-19 patient serum reveals potential prognostic markers of both severity and outcome. Metabolomics 18, 6 (2021). [PubMed: 34928464]
55. Danlos F-X et al. Metabolomic analyses of COVID-19 patients unravel stage-dependent and prognostic biomarkers. Cell Death Dis. 12, 258 (2021). [PubMed: 33707411]
56. Blasco H et al. The specific metabolome profiling of patients infected by SARS-COV-2 supports the key role of tryptophan-nicotinamide pathway and cytosine metabolism. Sci. Rep. 10, 16824 (2020). [PubMed: 33033346]
57. Thomas T et al. COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. JCI Insight 5, (2020).
58. Shen B et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. Cell 182, 59–72.e15 (2020). [PubMed: 32492406]
59. Barberis E et al. Large-Scale Plasma Analysis Revealed New Mechanisms and Molecules Associated with the Host Response to SARS-CoV-2. Int. J. Mol. Sci. 21, (2020).
60. Danchin A & Marlière P Cytosine drives evolution of SARS-CoV-2: Cytosine drives evolution of SARS-CoV-2. Environ. Microbiol. 22, 1977–1985 (2020). [PubMed: 32291894]
61. Kornhuber J, Hoertel N & Gulbins E The acid sphingomyelinase/ceramide system in COVID-19. Mol. Psychiatry 27, 307–314 (2022). [PubMed: 34608263]
62. Warburg O On the origin of cancer cells. Science 123, 309–314 (1956). [PubMed: 13298683]
63. Cori CF & Cori GT The carbohydrate metabolism of tumors. J. Biol. Chem. 64, 11–22 (1925).
64. Hui S et al. Glucose feeds the TCA cycle via circulating lactate. Nature 551, 115–118 (2017). [PubMed: 29045397]
65. Faubert B et al. Lactate Metabolism in Human Lung Tumors. Cell 171, 358–371.e9 (2017). [PubMed: 28985563]
66. DeBerardinis RJ & Chandel NS We need to talk about the Warburg effect. Nat Metab 2, 127–129 (2020). [PubMed: 32694689]
67. Chen P-H et al. Metabolic Diversity in Human Non-Small Cell Lung Cancer Cells. Mol. Cell 76, 838–851.e5 (2019). [PubMed: 31564558]
68. Santos CR & Schulze A Lipid metabolism in cancer. FEBS J. 279, 2610–2623 (2012). [PubMed: 22621751]

69. Huang J, Zhao B, Weinstein SJ, Albanes D & Mondul AM Metabolomic profile of prostate cancer-specific survival among 1812 Finnish men. BMC Med. 20, 362 (2022). [PubMed: 36280842]
70. Reichard CA et al. Gut Microbiome-Dependent Metabolic Pathways and Risk of Lethal Prostate Cancer: Prospective Analysis of a PLCO Cancer Screening Trial Cohort. Cancer Epidemiol. Biomarkers Prev. 31, 192–199 (2022). [PubMed: 34711629]
71. Richman EL et al. Choline intake and risk of lethal prostate cancer: incidence and survival. Am. J. Clin. Nutr. 96, 855–863 (2012). [PubMed: 22952174]
72. Bae S et al. Plasma choline metabolites and colorectal cancer risk in the Women's Health Initiative Observational Study. Cancer Res. 74, 7442–7452 (2014). [PubMed: 25336191]
73. Huang J et al. Serum metabolomic profiling of prostate cancer risk in the prostate, lung, colorectal, and ovarian cancer screening trial. Br. J. Cancer 115, 1087–1095 (2016). [PubMed: 27673363]
74. Mondul AM et al. Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocolpherol, beta-carotene cancer prevention (ATBC) study. Int. J. Cancer 137, 2124–2132 (2015). [PubMed: 25904191]
75. Schmidt JA et al. Patterns in metabolite profile are associated with risk of more aggressive prostate cancer: A prospective study of 3,057 matched case-control sets from EPIC. Int. J. Cancer 146, 720–730 (2020). [PubMed: 30951192]
76. Shu X et al. Prospective study of blood metabolites associated with colorectal cancer risk. Int. J. Cancer 143, 527–534 (2018). [PubMed: 29479691]
77. Kühn T et al. Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study. BMC Med. 14, 13 (2016). [PubMed: 26817443]
78. Lu M-S et al. Choline and betaine intake and colorectal cancer risk in Chinese population: a case-control study. PLoS One 10, e0118661 (2015). [PubMed: 25785727]
79. Zhang C-X et al. Choline and betaine intake is inversely associated with breast cancer risk: a two-stage case-control study in China. Cancer Sci. 104, 250–258 (2013). [PubMed: 23140534]
80. Zeng F-F et al. Choline and betaine intakes are associated with reduced risk of nasopharyngeal carcinoma in adults: a case-control study. Br. J. Cancer 110, 808–816 (2014). [PubMed: 24169354]
81. Sun S et al. Choline and betaine consumption lowers cancer risk: a meta-analysis of epidemiologic studies. Sci. Rep. 6, 35547 (2016). [PubMed: 27759060]
82. Goldberg AA, Titorenko VI, Beach A & Sanderson JT Bile acids induce apoptosis selectively in androgen-dependent and -independent prostate cancer cells. PeerJ 1, e122 (2013). [PubMed: 23940835]
83. Stepien M et al. Metabolic perturbations prior to hepatocellular carcinoma diagnosis: Findings from a prospective observational cohort study. Int. J. Cancer 148, 609–625 (2021). [PubMed: 32734650]
84. Stepien M et al. Prediagnostic alterations in circulating bile acid profiles in the development of hepatocellular carcinoma. Int. J. Cancer 150, 1255–1268 (2022). [PubMed: 34843121]
85. Thomas CE et al. Association between Pre-Diagnostic Serum Bile Acids and Hepatocellular Carcinoma: The Singapore Chinese Health Study. Cancers 13, (2021).
86. Yuan J-M, Gao Y-T, Ong C-N, Ross RK & Yu MC Prediagnostic level of serum retinol in relation to reduced risk of hepatocellular carcinoma. J. Natl. Cancer Inst. 98, 482–490 (2006). [PubMed: 16595784]
87. Wang Y, Jacobs EJ, Carter BD, Gapstur SM & Stevens VL Plasma Metabolomic Profiles and Risk of Advanced and Fatal Prostate Cancer. Eur Urol Oncol 4, 56–65 (2021). [PubMed: 31378665]
88. Meller S et al. Integration of tissue metabolomics, transcriptomics and immunohistochemistry reveals ERG- and gleason score-specific metabolomic alterations in prostate cancer. Oncotarget 7, 1421–1438 (2016). [PubMed: 26623558]
89. Dudka I et al. Comprehensive metabolomics analysis of prostate cancer tissue in relation to tumor aggressiveness and TMPRSS2-ERG fusion status. BMC Cancer 20, 437 (2020). [PubMed: 32423389]

90. Feng X et al. Association of Prediagnostic Blood Metabolomics with Prostate Cancer Defined by ERG or PTEN Molecular Subtypes. Cancer Epidemiol. Biomarkers Prev. 30, 1000–1008 (2021). [PubMed: 33627383]
91. Xiao Y et al. Comprehensive metabolomics expands precision medicine for triple-negative breast cancer. Cell Res. 32, 477–490 (2022). [PubMed: 35105939]
92. Jack CR Jr et al. Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease. Brain 132, 1355–1365 (2009). [PubMed: 19339253]
93. Toledo JB et al. Metabolic network failures in Alzheimer's disease: A biochemical road map. Alzheimers. Dement. 13, 965–984 (2017). [PubMed: 28341160]
94. Varma VR et al. Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. PLoS Med. 15, e1002482 (2018). [PubMed: 29370177]
95. Wang G et al. Plasma metabolite profiles of Alzheimer's disease and mild cognitive impairment. J. Proteome Res. 13, 2649–2658 (2014). [PubMed: 24694177]
96. Serrano-Pozo A, Frosch MP, Masliah E & Hyman BT Neuropathological alterations in Alzheimer disease. Cold Spring Harb. Perspect. Med. 1, a006189 (2011). [PubMed: 22229116]
97. Dong R et al. CSF metabolites associate with CSF tau and improve prediction of Alzheimer's disease status. Alzheimers. Dement. 13, e12167 (2021).
98. Darst BF, Lu Q, Johnson SC & Engelman CD Integrated analysis of genomics, longitudinal metabolomics, and Alzheimer's risk factors among 1,111 cohort participants. Genet. Epidemiol. 43, 657–674 (2019). [PubMed: 31104335]
99. Wörheide MA et al. An integrated molecular Atlas of Alzheimer's disease. bioRxiv (2021) doi:10.1101/2021.09.14.21263565.
100. Lloyd-Price J et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569, 655–662 (2019). [PubMed: 31142855]
101. Nikolaus S et al. Increased Tryptophan Metabolism Is Associated With Activity of Inflammatory Bowel Diseases. Gastroenterology 153, 1504–1516.e2 (2017). [PubMed: 28827067]
102. Scoville EA et al. Alterations in Lipid, Amino Acid, and Energy Metabolism Distinguish Crohn's Disease from Ulcerative Colitis and Control Subjects by Serum Metabolomic Profiling. Metabolomics 14, 17 (2018). [PubMed: 29681789]
103. Dawiskiba T et al. Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases. World J. Gastroenterol. 20, 163–174 (2014). [PubMed: 24415869]
104. Gallagher K, Catesson A, Griffin JL, Holmes E & Williams HRT Metabolomic Analysis in Inflammatory Bowel Disease: A Systematic Review. J. Crohns. Colitis 15, 813–826 (2021). [PubMed: 33175138]
105. Hua X et al. Inflammatory Bowel Disease Is Associated With Prediagnostic Perturbances in Metabolic Pathways. Gastroenterology 164, 147–150.e2 (2023). [PubMed: 36122699]
106. Coresh J et al. Calibration and random variation of the serum creatinine assay as critical elements of using equations to estimate glomerular filtration rate. Am. J. Kidney Dis. 39, 920–929 (2002). [PubMed: 11979335]
107. Peng H et al. Identification of Metabolite Markers Associated with Kidney Function. J Immunol Res 2022, 6190333 (2022). [PubMed: 35928631]
108. Yu B et al. Serum metabolomic profiling and incident CKD among African Americans. Clin. J. Am. Soc. Nephrol. 9, 1410–1417 (2014). [PubMed: 25011442]
109. Chen D-Q et al. Identification of serum metabolites associating with chronic kidney disease progression and anti-fibrotic effect of 5-methoxytryptophan. Nat. Commun. 10, 1476 (2019). [PubMed: 30931940]
110. Fonseca RID et al. Untargeted plasma 1H NMR-based metabolomic profiling in different stages of chronic kidney disease. J. Pharm. Biomed. Anal. 229, 115339 (2023). [PubMed: 36963247]
111. Zhang F et al. Untargeted serum metabolomics and tryptophan metabolism profiling in type 2 diabetic patients with diabetic glomerulopathy. Ren. Fail. 43, 980–992 (2021). [PubMed: 34157945]

112. McClellan WM & Flanders WD Risk factors for progressive chronic kidney disease. J. Am. Soc. Nephrol. 14, S65–70 (2003). [PubMed: 12819305]
113. Hu D-Y et al. Metabolomics analysis of human plasma reveals decreased production of trimethylamine N-oxide retards the progression of chronic kidney disease. Br. J. Pharmacol. 179, 4344–4359 (2022). [PubMed: 35428974]
114. Goek O-N et al. Serum metabolite concentrations and decreased GFR in the general population. Am. J. Kidney Dis. 60, 197–206 (2012). [PubMed: 22464876]
115. Benito S et al. Untargeted metabolomics for plasma biomarker discovery for early chronic kidney disease diagnosis in pediatric patients using LC-QTOF-MS. Analyst 143, 4448–4458 (2018). [PubMed: 30151522]
116. Fuller H, Iles M, Moore JB & Zulyniak MA Unique Metabolic Profiles Associate with Gestational Diabetes and Ethnicity in Low- and High-Risk Women Living in the UK. J. Nutr. 152, 2186–2197 (2022). [PubMed: 35883228]
117. Nobakht M Gh BF Application of metabolomics to preeclampsia diagnosis. Syst. Biol. Reprod. Med. 64, 324–339 (2018). [PubMed: 29965778]
118. Ciborowski M et al. Potential first trimester metabolomic biomarkers of abnormal birth weight in healthy pregnancies. Prenat. Diagn. 34, 870–877 (2014). [PubMed: 24733416]
119. Bot M et al. Metabolomics Profile in Depression: A Pooled Analysis of 230 Metabolic Markers in 5283 Cases With Depression and 10,145 Controls. Biol. Psychiatry 87, 409–418 (2020). [PubMed: 31635762]
120. MacDonald K et al. Biomarkers for major depressive and bipolar disorders using metabolomics: A systematic review. Am. J. Med. Genet. B Neuropsychiatr. Genet. 180, 122–137 (2019). [PubMed: 30411484]
121. Zhu Y et al. Psychological distress and metabolomic markers: A systematic review of posttraumatic stress disorder, anxiety, and subclinical distress. Neurosci. Biobehav. Rev. 143, 104954 (2022). [PubMed: 36368524]
122. Winning A, Glymour MM, McCormick MC, Gilsanz P & Kubzansky LD Psychological Distress Across the Life Course and Cardiometabolic Risk: Findings From the 1958 British Birth Cohort Study. J. Am. Coll. Cardiol. 66, 1577–1586 (2015). [PubMed: 26429083]
123. Hagenbeek FA et al. Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. Nat. Commun. 11, 39 (2020). [PubMed: 31911595]
124. Yousri NA et al. Whole-exome sequencing identifies common and rare variant metabolic QTLs in a Middle Eastern population. Nat. Commun. 9, 333 (2018). [PubMed: 29362361]
125. Darst BF, Koscik RL, Hogan KJ, Johnson SC & Engelman CD Longitudinal plasma metabolomics of aging and sex. Aging 11, 1262–1282 (2019). [PubMed: 30799310]
126. Chen Y et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. Nat. Genet. 55, 44–53 (2023). [PubMed: 36635386]
127. Speed D & Balding DJ SumHer better estimates the SNP heritability of complex traits from summary statistics. Nat. Genet. 51, 277–284 (2019). [PubMed: 30510236]
128. Wright FA et al. Heritability and genomics of gene expression in peripheral blood. Nat. Genet. 46, 430–437 (2014). [PubMed: 24728292]
129. Lloyd-Jones LR et al. The Genetic Architecture of Gene Expression in Peripheral Blood. Am. J. Hum. Genet. 100, 228–237 (2017). [PubMed: 28065468]
130. Ouwens KG et al. A characterization of cis- and trans-heritability of RNA-Seq-based gene expression. Eur. J. Hum. Genet. 28, 253–263 (2020). [PubMed: 31558840]
131. Smith CJ et al. Integrative analysis of metabolite GWAS illuminates the molecular basis of pleiotropy and genetic correlation. Elife 11, (2022).
132. Yin X et al. Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. Nat. Commun. 13, 1644 (2022). [PubMed: 35347128]
133. Tahir UA et al. Whole Genome Association Study of the Plasma Metabolome Identifies Metabolites Linked to Cardiometabolic Disease in Black Individuals. Nat. Commun. 13, 4923 (2022). [PubMed: 35995766]

134. Harshfield EL et al. Genome-wide analysis of blood lipid metabolites in over 5000 South Asians reveals biological insights at cardiometabolic disease loci. BMC Med. 19, 232 (2021). [PubMed: 34503513]
135. Yu B et al. Genetic determinants influencing human serum metabolome among African Americans. PLoS Genet. 10, e1004212 (2014). [PubMed: 24625756]
136. Feofanova EV et al. A Genome-wide Association Study Discovers 46 Loci of the Human Metabolome in the Hispanic Community Health Study/Study of Latinos. Am. J. Hum. Genet. 107, 849–863 (2020). [PubMed: 33031748]
137. Wang Z et al. Genome-wide association study of metabolites in patients with coronary artery disease identified novel metabolite quantitative trait loci. Clin. Transl. Med. 11, e290 (2021). [PubMed: 33634981]
138. Jung T et al. Integrative Pathway Analysis of SNP and Metabolite Data Using a Hierarchical Structural Component Model. Front. Genet. 13, 814412 (2022). [PubMed: 35401680]
139. Fuller H, Iles MM, Moore JB & Zulyniak MA Metabolic drivers of dysglycemia in pregnancy: ethnic-specific GWAS of 146 metabolites and 1-sample Mendelian randomization analyses in a UK multi-ethnic birth cohort. Front. Endocrinol. 14, 1157416 (2023).
140. Vriens K et al. Evidence for an alternative fatty acid desaturation pathway increasing cancer plasticity. Nature 566, 403–406 (2019). [PubMed: 30728499]
141. Yuan S et al. Plasma Phospholipid Fatty Acids, FADS1 and Risk of 15 Cardiovascular Diseases: A Mendelian Randomisation Study. Nutrients 11, (2019).
142. Brayner B, Kaur G, Keske MA & Livingstone KM FADS Polymorphism, Omega-3 Fatty Acids and Diabetes Risk: A Systematic Review. Nutrients 10, (2018).
143. Mathias RA, Pani V & Chilton FH Genetic Variants in the FADS Gene: Implications for Dietary Recommendations for Fatty Acid Intake. Curr. Nutr. Rep. 3, 139–148 (2014). [PubMed: 24977108]
144. Sollis E et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 51, D977–D985 (2023). [PubMed: 36350656]
145. Chang L, Zhou G, Ou H & Xia J mGWAS-Explorer: Linking SNPs, Genes, Metabolites, and Diseases for Functional Insights. Metabolites 12, (2022).
146. Staley JR et al. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics 32, 3207–3209 (2016). [PubMed: 27318201]
147. Gagliano Taliun SA et al. Exploring and visualizing large-scale genetic associations by using PheWeb. Nat. Genet. 52, 550–552 (2020). [PubMed: 32504056]
148. Fang S, Holmes MV, Gaunt TR, Davey Smith G & Richardson TG Constructing an atlas of associations between polygenic scores from across the human phenome and circulating metabolic biomarkers. Elife 11, (2022).
149. Manor O et al. Health and disease markers correlate with gut microbiome composition across thousands of people. Nat. Commun. 11, 5206 (2020). [PubMed: 33060586]
150. Zierer J et al. The fecal metabolome as a functional readout of the gut microbiome. Nat. Genet. 50, 790–795 (2018). [PubMed: 29808030]
151. Wang M et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. 34, 828–837 (2016). [PubMed: 27504778]
152. Quinn RA et al. Global chemical effects of the microbiome include new bile-acid conjugations. Nature 579, 123–129 (2020). [PubMed: 32103176]
153. Folz JS, Shalon D & Fiehn O Metabolomics analysis of time-series human small intestine lumen samples collected in vivo. Food Funct. 12, 9405–9415 (2021). [PubMed: 34606553]
154. Patel CJ, Bhattacharya J & Butte AJ An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. PLoS One 5, e10746 (2010). [PubMed: 20505766]
155. Price EJ et al. Merging the exposome into an integrated framework for ‘omics’ sciences. iScience 25, 103976 (2022). [PubMed: 35310334]

156. Huang S-Y et al. Investigating Causal Relations Between Circulating Metabolites and Alzheimer's Disease: A Mendelian Randomization Study. J. Alzheimers. Dis. 87, 463–477 (2022). [PubMed: 35275550]

157. Yu X-H, Cao R-R, Yang Y-Q & Lei S-F Identification of causal metabolites related to multiple autoimmune diseases. Hum. Mol. Genet. 31, 604–613 (2022). [PubMed: 34523675]

158. Porcu E et al. Triangulating evidence from longitudinal and Mendelian randomization studies of metabolomic biomarkers for type 2 diabetes. Sci. Rep. 11, 6197 (2021). [PubMed: 33737653]

159. Feng Y et al. Causal effects of genetically determined metabolites on cancers included lung, breast, ovarian cancer, and glioma: a Mendelian randomization study. Transl Lung Cancer Res 11, 1302–1314 (2022). [PubMed: 35958335]

160. Dreyfuss JM et al. High-throughput mediation analysis of human proteome and metabolome identifies mediators of post-bariatric surgical diabetes control. Nat. Commun. 12, 6951 (2021). [PubMed: 34845204]

161. Camacho D, de la Fuente A & Mendes P The origin of correlations in metabolomics data. Metabolomics 1, 53–63 (2005).

162. Kelly RS et al. An Integrative Transcriptomic and Metabolomic Study of Lung Function in Children With Asthma. Chest 154, 335–348 (2018). [PubMed: 29908154]

163. Krumsiek J, Suhre K, Illig T, Adamski J & Theis FJ Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. BMC Syst. Biol. 5, 21 (2011). [PubMed: 21281499]

164. Krumsiek J et al. Gender-specific pathway differences in the human serum metabolome. Metabolomics 11, 1815–1833 (2015). [PubMed: 26491425]

165. Playdon MC et al. Nutritional metabolomics and breast cancer risk in a prospective study. Am. J. Clin. Nutr. 106, 637–649 (2017). [PubMed: 28659298]

166. Lawlor DA, Tilling K & Davey Smith G Triangulation in aetiological epidemiology. Int. J. Epidemiol. 45, 1866–1886 (2016). [PubMed: 28108528]

167. Nemet I et al. A Cardiovascular Disease-Linked Gut Microbial Metabolite Acts via Adrenergic Receptors. Cell 180, 862–877.e22 (2020). [PubMed: 32142679]

168. Adam J et al. Metformin Effect on Nontargeted Metabolite Profiles in Patients With Type 2 Diabetes and in Multiple Murine Tissues. Diabetes 65, 3776–3785 (2016). [PubMed: 27621107]

169. Wu ZE et al. Metabolomic signatures for visceral adiposity and dysglycaemia in Asian Chinese and Caucasian European adults: the cross-sectional TOFI_Asia study. Nutr. Metab. 17, 95 (2020).

170. Taylor K et al. Differences in Pregnancy Metabolic Profiles and Their Determinants between White European and South Asian Women: Findings from the Born in Bradford Cohort. Metabolites 9, (2019).

171. Chevli PA et al. Plasma metabolomic profiling in subclinical atherosclerosis: the Diabetes Heart Study. Cardiovasc. Diabetol. 20, 231 (2021). [PubMed: 34876126]

172. Saravanan P, Diabetes in Pregnancy Working Group, Maternal Medicine Clinical Study Group & Royal College of Obstetricians and Gynaecologists, UK. Gestational diabetes: opportunities for improving maternal and child health. Lancet Diabetes Endocrinol 8, 793–800 (2020). [PubMed: 32822601]

173. Geidenstam N et al. Using metabolite profiling to construct and validate a metabolite risk score for predicting future weight gain. PLoS One 14, e0222445 (2019). [PubMed: 31560688]

174. Li C et al. Development and validation of a metabolite score for red meat intake: an observational cohort study and randomized controlled dietary intervention. Am. J. Clin. Nutr. 116, 511–522 (2022). [PubMed: 35754192]

175. Li J et al. The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. Eur. Heart J. 41, 2645–2656 (2020). [PubMed: 32406924]

176. Fernández-Ochoa Á et al. Metabolic Disturbances in Urinary and Plasma Samples from Seven Different Systemic Autoimmune Diseases Detected by HPLC-ESI-QTOF-MS. J. Proteome Res. 19, 3220–3229 (2020). [PubMed: 32460496]

177. Liu N et al. Comparison of Untargeted Metabolomic Profiling vs Traditional Metabolic Screening to Identify Inborn Errors of Metabolism. JAMA Netw Open 4, e2114155 (2021). [PubMed: 34251446]

178. Goossens N, Nakagawa S, Sun X & Hoshida Y Cancer biomarker discovery and validation. Transl. Cancer Res. 4, 256–269 (2015). [PubMed: 26213686]

179. Braisted J et al. RaMP-DB 2.0: a renovated knowledgebase for deriving biological and chemical insight from metabolites, proteins, and genes. Bioinformatics 39, (2023).

180. Wang Z et al. Genome-wide association study of metabolites in patients with coronary artery disease identified novel metabolite quantitative trait loci. Clin. Transl. Med. 11, e290 (2021). [PubMed: 33634981]

181. Jung T et al. Integrative Pathway Analysis of SNP and Metabolite Data Using a Hierarchical Structural Component Model. Front. Genet. 13, 814412 (2022). [PubMed: 35401680]

182. Wolfe D, Dudek S, Ritchie MD & Pendergrass SA Visualizing genomic information across chromosomes with PhenoGram. BioData Min. 6, 18 (2013). [PubMed: 24131735]

183. VanderWeele TJ & VanderWeele T Explanation in Causal Inference: Methods for Mediation and Interaction. (Oxford University Press, USA, 2015).

184. Hemani G, Tilling K & Davey Smith G Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. 13, e1007081 (2017). [PubMed: 29149188]

185. Verbanck M, Chen C-Y, Neale B & Do R Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat. Genet. 50, 693–698 (2018). [PubMed: 29686387]

186. Burgess S & Thompson SG Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am. J. Epidemiol. 181, 251–260 (2015). [PubMed: 25632051]

187. Davey Smith G & Hemani G Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum. Mol. Genet. 23, R89–98 (2014). [PubMed: 25064373]

188. Zhang B & Horvath S A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4, Article17 (2005).

189. Shutta KH, De Vito R, Scholtens DM & Balasubramanian R Gaussian graphical models with applications to omics analyses. Stat. Med. 41, 5150–5187 (2022). [PubMed: 36161666]

190. Dettori JR, Norvell DC & Chapman JR Fixed-Effect vs Random-Effects Models for Meta-Analysis: 3 Points to Consider. Global Spine J 12, 1624–1626 (2022). [PubMed: 35723546]

191. Fitzmaurice GM, Laird NM & Ware JH Applied Longitudinal Analysis. (John Wiley & Sons, 2011).

Box 1.

Metabolomic epidemiology terminologies.

Epidemiology: The study of the distribution and determinants of health-related states and events in populations.

Metabolomics: The unbiased identification and quantification of all metabolites in a biological system.

Metabolomic epidemiology: The high-throughput study of the relationship between metabolites and health-related traits in population-based epidemiologic studies.

Targeted metabolomics: The analysis of a pre-defined subset of metabolites (often hypothesis driven) that are chemically characterised and biochemically annotated.

Untargeted metabolomics: Hypothesis generating top-down strategy that analyse all measurable metabolites (within the range of the platform used) in a biological sample simultaneously.

Mass spectrometry (MS): Analytical technique that measures the mass/charge (m/z) ratio to identify metabolites within a sample.

Nuclear magnetic resonance (NMR) spectroscopy: Detects metabolites by determining the molecular structure of the molecule from the electromagnetic spectrum absorbed and readmitted when the sample is exposed to a magnetic field.

Absolute quantification: Determines the levels of absolute abundance of a metabolite in a sample using a standard curve method. An internal standard may be required, depending on the analytical method used.

Relative quantification: Determines metabolite spectral patterns and intensities in a sample relative to a reference sample.

Unknown or unidentified metabolites: Molecules that have been detected in a sample, but the molecular structure has not been determined.

Endogenous metabolites: Metabolites that are largely naturally produced by an organism and influenced by the genome, epigenome, transcriptome, and proteome.

Exogenous metabolites: Metabolites that are largely produced by external factors, such as diet, medications, environmental exposures, and lifestyle.

Box 2.

Statistical Methods: Causal Mediation Analysis, Mendelian Randomization, and Correlation Networks.

Causal Mediation Analysis

Mediation analysis models the causal mechanism underlying the relationship between an exposure and outcome by disentangling the direct effect of the exposure on the outcome from the indirect effect mediated by a set of other variables ("mediators"). In recent years, the counterfactual approach to mediation analysis has allowed for more rigorous definitions of causal effects under explicitly stated identifiability assumptions and the development of more advanced models including multiple mediators and interactions[183].

Under the counterfactual framework, the total effect between the exposure and outcome can be decomposed into the natural direct effect (NDE) and the natural indirect effect (NIE). The NDE is the effect when the mediator takes the value it would naturally take in the absence of the exposure. The NIE is the effect of switching the mediator from the value it would have taken in the absence versus presence of the exposure. To identify the NDE and NIE, it is required that 1) there be no unmeasured common causes of the exposure, mediator, and outcome and 2) no mediator-outcome confounders are affected by the exposure. In the setting of metabolomic epidemiology, if metabolites are considered as mediators, technical and biological variables that may not seem to confound the exposure-outcome relationship become crucial. Importantly, when multiple metabolites are involved in the causal mechanism, it may not be possible to disentangle every metabolite-specific pathway due to dependence between the metabolite levels.

Various estimation methods are available for causal mediation analysis and implemented in common software packages (e.g., CMAverse and mediation packages in R and PROC CAUSALMED in SAS).

Mendelian Randomisation

Mendelian Randomisation (MR) is a statistical technique used to evaluate potentially causal associations between an exposure and outcome. A common approach to conducting MR uses summary statistics from a GWAS on the 1) exposure of interest and 2) the outcome of interest, referred to as two-sample MR[184]. For unbiased estimates, these GWAS should be conducted in non-overlapping participants from the same population.

The validity of MR is based on three assumptions: 1) IVs must be associated with the exposure, 2) must only associate with the outcome through the pathway in question (exclusion restriction), and 3) should be independent of confounders impacting both the exposure and outcome (i.e., no horizontal pleiotropy). Correlation between metabolites may violate the second and third assumption; therefore, MR findings should be interpreted with caution in the context of metabolomic epidemiology. MR sensitivity analyses can evaluate these assumptions and provide consistent effect estimates in the presence of pleiotropy. For instance, MR-PRESSO identifies and removes variants in horizontal pleiotropy while providing an estimate of the extent to which pleiotropic

variants impact results[185]. MR has also been extended to multivariable MR, which jointly models multiple exposures[186] and to model bidirectional designs where the exposure and outcome are reversed in a secondary analysis[187].

While MR can provide insights into potentially causal mechanisms between an exposure and outcome, it cannot unravel more complex relationships between genetic and other factors. Using MR in conjunction with other statistical and experimental approaches can help address limitations of this approach.

Correlation and partial correlation networks

Correlation networks capture the correlation structure of data in a network model, which consists of nodes representing metabolites and edges representing the correlation between them. Weighted Gene Coexpression Network Analysis (WGCNA) is an extension of this concept in which weighted edges are estimated based on soft thresholding of the between-metabolite correlations[188]. WGCNA can also be used to identify network modules and calculate module eigengenes, which can then be used to assign summary module scores.

Gaussian graphical models (GGMs, or partial correlation networks) are network models in which nodes correspond to variables (metabolites) and edges between two metabolites represent their partial correlation, i.e., their correlation conditioned on the other metabolites in the network[189]. These partial correlations represent direct pairwise associations between metabolites that cannot be explained by any of the other metabolites in the network; as such, GGMs are typically sparser than correlation networks and edges in GGMs reflect more direct associations.

Box 3.

Statistical Methods: Meta-Analysis and Longitudinal Analysis.

Meta-analysis

Meta-analysis combines data from multiple studies to obtain a single estimate of an exposure-outcome association. Meta-analysis is particularly useful in the field of metabolomic epidemiology, where individual studies tend to be small and may be under-powered. The most common methods for meta-analyses are fixed effect and random effect models. Fixed effect models assume that there is one true effect size across studies, with intra-study variability being solely a result of sampling variability. Random effect models assume heterogeneity of the true effect size across studies resulting from both intra- and inter-study variability[190]. Effect estimates resulting from meta-analysis studies are weighted based on sample size or error variance, giving more weight to studies with larger sample sizes or smaller error variance. In metabolomic epidemiology, care should be taken to ensure that metabolites measured are adequately harmonized and comparable between studies prior to analysis so that metabolites are similar enough to be combined.

Longitudinal analysis

Longitudinal analysis can provide insight into the metabolic changes that coincide with disease progression and partially account for temporal variation in metabolite measures from confounding factors such as diurnal variation and postprandial responses. Mixed models for longitudinal analysis explicitly account for the dependence structure of repeated measures. These models partition the exposure-response relationship into fixed effects, which model the average relationship between exposure and response among the population, and random effects, which model the deviation of an individual's exposure-response relationship from the population average[191]. There are several well-established software packages for mixed model estimation, including the lme4 package in R and PROC MIXED in SAS.

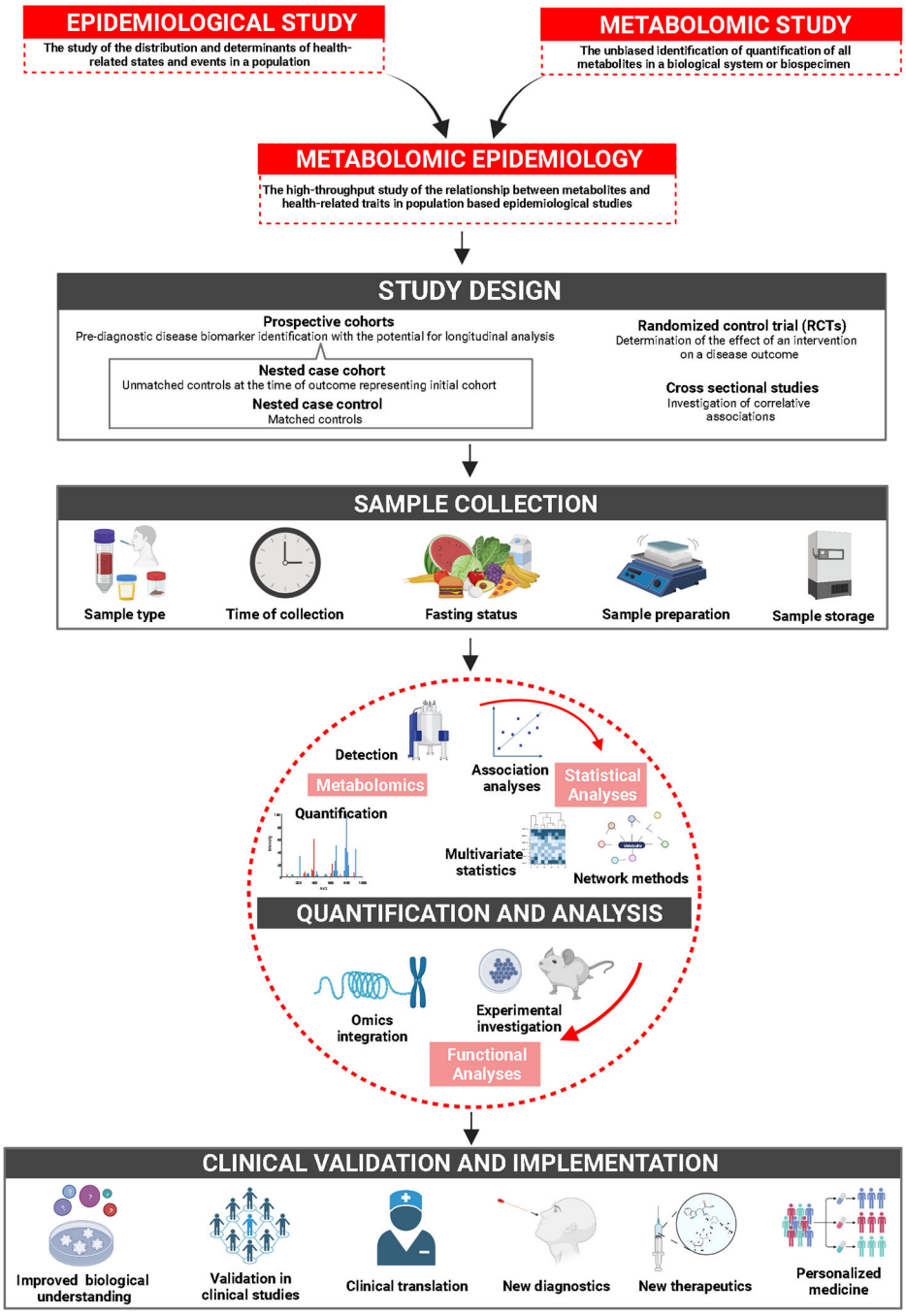

Figure 1.
Summary of the process and applications of metabolomic epidemiology studies.

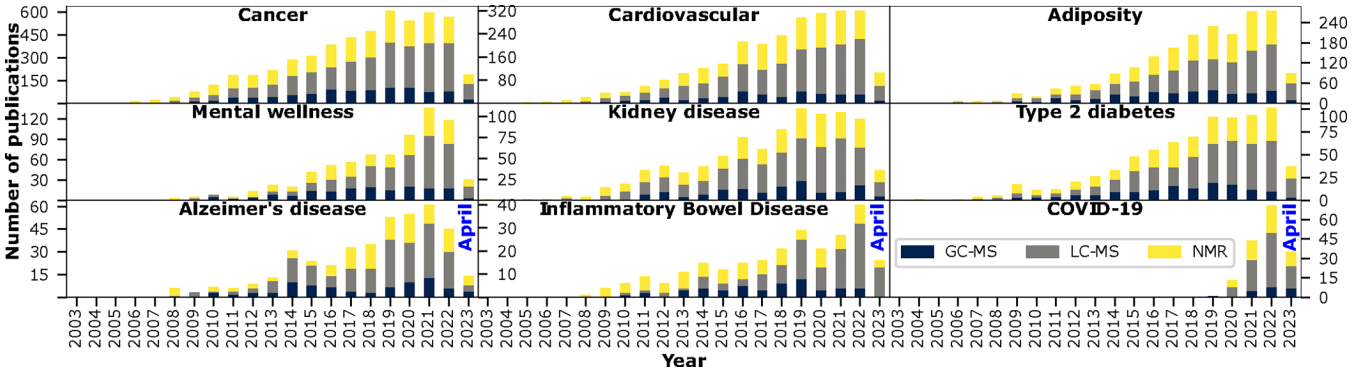

Figure 2. Number of metabolomic publications across commonly studied health-related traits over the years.

Results are based on a PubMed database search conducted on April 27, 2023. Search terms included the condition of interest along with metabolom* and NMR, GC-MS, or LC-MS. Each search was conducted separately; as such, it is possible that studies using more than one technology were counted twice. Search terms used for “mental health” included mental health, mental wellness, psychological distress, depression, anxiety, posttraumatic stress disorder, or PTSD. Search terms for “adiposity” included adiposity, BMI, body-mass index, WHR, waist-hip ratio, or obesity. Search terms for “COVID-19” included COVID-19, 2019-nCoV, SARS-CoV-2, Coronavirus-2, or coronavirus 19. Search terms for “Cardiovascular” included cardiovascular, stroke, heart failure, coronary artery disease, coronary heart disease, venous thromboembolism, pulmonary embolism, myocardial infarction, cardiovascular mortality, or major cardiovascular event.

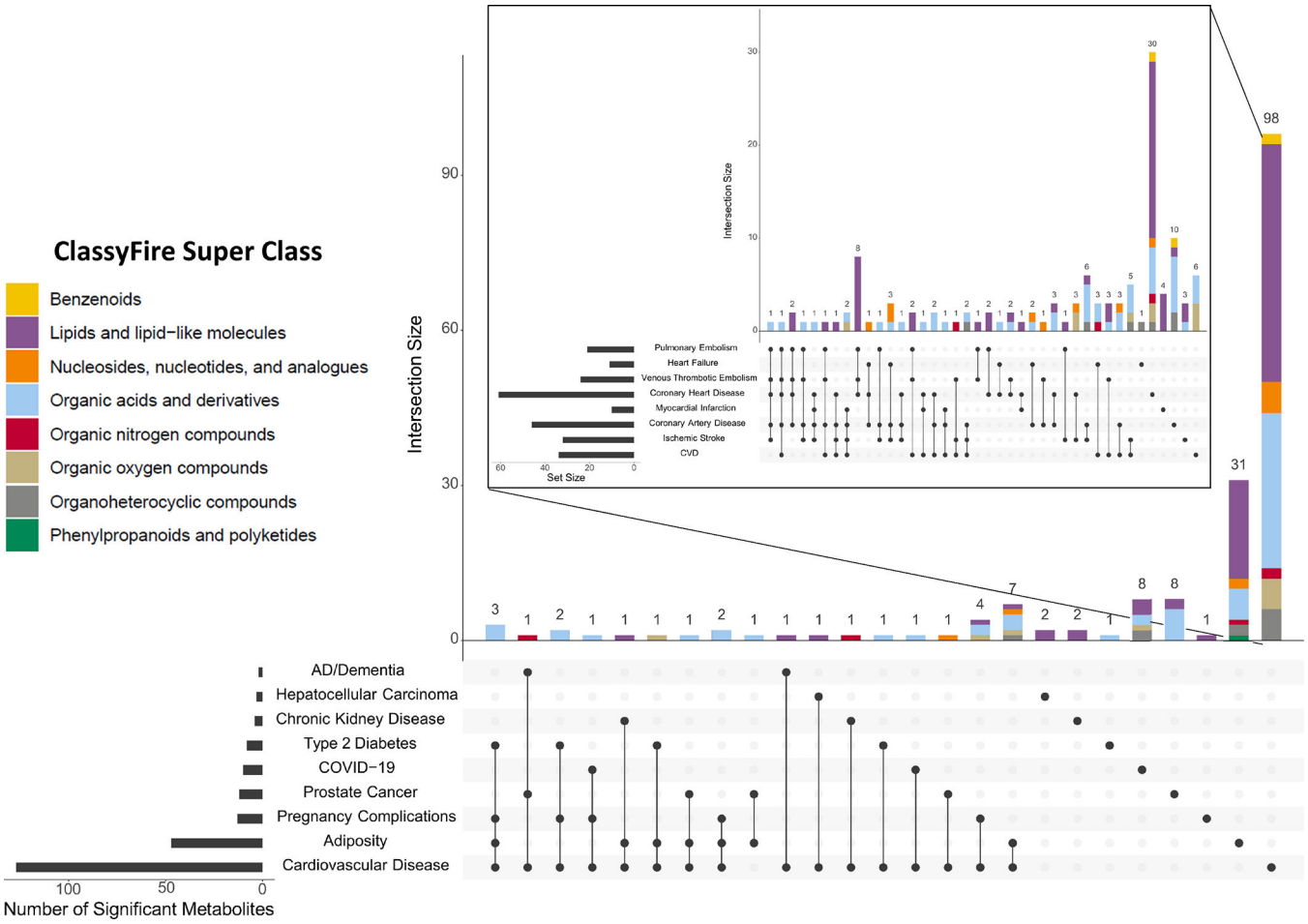

Figure 3. Summary of consistently reported circulating metabolites associated with traits discussed in this review.

Metabolites reported are significantly associated with the trait of interest (FDR adjusted p-values <0.05) in a metabolomic investigation, with replication in at least one independent study. Due to the large number of metabolomic studies conducted in cardiovascular disease (CVD), only CVD studies exceeding >100 participants were included for the purpose of creating this figure. These UpSet plots indicate the number of metabolites uniquely associated with one indicated trait (e.g., 31 are associated with adiposity and no other traits shown here) and the number of metabolites associated with multiple traits (e.g., seven metabolites are associated with both adiposity and CVD). The inset plot depicts metabolites associated with sub-conditions of CVD. ClassyFire super classes were assigned to metabolites using RaMP-DB 2.0[179].

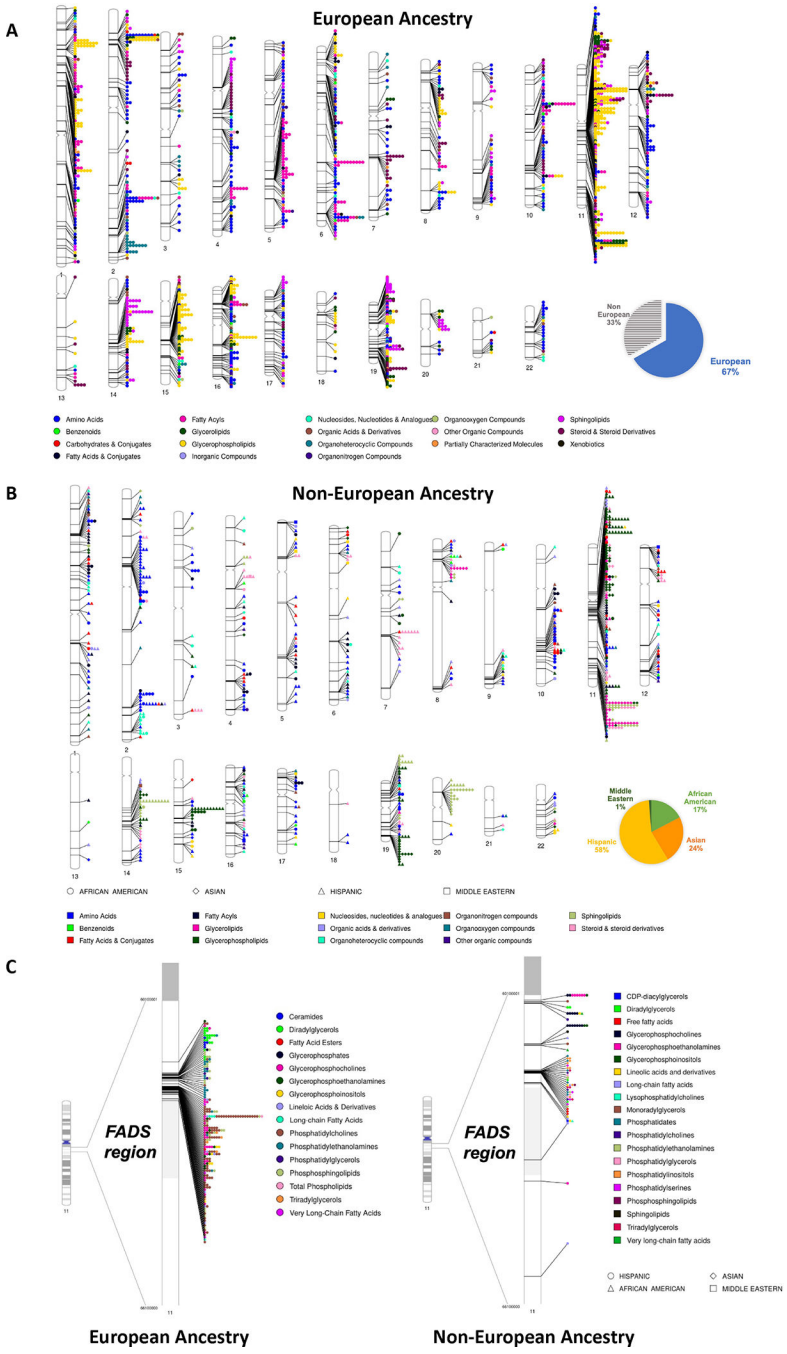

Figure 4. Significant associations identified in circulating metabolomic GWAS.

Studies were identified by conducting a literature review, including the studies cited in Supplementary Table 8 of Chen et al (2018)[124,133–136,139,180,181]. When the same metabolite was identified more than once within a 500kb region, the strongest single association was included. **A**. Significant associations identified in individuals of European ancestry. **B.** Significant associations identified in individuals with significant non-European ancestry. Pie charts in A and B indicate the proportion of significant associations identified by population. **C.** Lipid metabolite associations identified in the FADS region. “Other Organic

Compounds" class includes prenol lipids, alkaloids and derivatives, phenylpropanoic acids, hydroxycinnamic acids, and peptides. Pie charts represent the number of significant metabolite-variant associations identified in each population, with associations based on the metabolomic GWAS references provided in the "Metabolomics and Genetics" section. Note that lipids and amino acids are much more commonly measured and numerous than other metabolite classes. Figure 4 was created with PhenoGram[182].