ANNUAL REVIEWS

Annual Review of Medicine

Omics and Cardiometabolic Disease Risk Prediction

Usman A. Tahir and Robert E. Gerszten

Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA; email: utahir@bidmc.harvard.edu, rgerszte@bidmc.harvard.edu

Annu. Rev. Med. 2020. 71:163–75

The *Annual Review of Medicine* is online at med.annualreviews.org

https://doi.org/10.1146/annurev-med-042418-010924

Keywords

omics, cardiometabolic risk, risk prediction, genomics, metabolomics, proteomics, multi-omics, multi-marker

Abstract

Risk assessments are integral for the prevention and management of cardiometabolic disease (CMD). However, individuals may develop CMD without traditional risk factors, necessitating the development of novel biomarkers to aid risk prediction. The emergence of omic technologies, including genomics, proteomics, and metabolomics, has allowed for assessment of orthogonal measures of cardiometabolic risk, potentially improving the ability for novel biomarkers to refine disease risk assessments. While omics has shed light on novel mechanisms for the development of CMD, its adoption in clinical practice faces significant challenges. We review select omic technologies and cardiometabolic investigations for risk prediction, while highlighting challenges and opportunities for translating findings to clinical practice.

INTRODUCTION

Cardiometabolic disease (CMD) is among the leading causes of morbidity and mortality worldwide and is a significant burden on healthcare resources. Risk assessments for CMD are an integral component of medical practice, providing clinicians with tools to evaluate preventative and therapeutic options in patients at greatest risk for developing disease. Seminal risk factors for cardiovascular disease (CVD) were identified over half a century ago (1), and while they are powerful tools, they are imperfect in their predictive abilities. The concept of residual risk has been increasingly appreciated, warranting research into the identification of novel biomarkers to bridge the gap and improve our understanding of disease processes. However, only a handful of biomarkers have made it to clinical practice for risk prediction, highlighting the inherent challenges in biomarker development.

Traditional risk factors for CMD perform quite well in determining disease risk, setting a reasonably high bar for the ability of new markers to improve risk prediction and necessitating biomarker discovery from novel disease mechanisms rather than established pathways. Recently, major technological advances in the field of omics—the broad-scale study of cellular molecules such as genes, proteins, and metabolites—have expanded our capacity to systematically interrogate large populations to discover novel disease risk factors. While the explosion of data has revealed new insights into disease pathophysiology, the utility of new molecular markers for CMD risk prediction in clinical practice remains to be rigorously evaluated (2, 3). In this review, we highlight select CMD omic investigations, novel predictors of disease, challenges in translating findings to clinical practice, and future directions for research. While the range of CMD is considerable, the focus of this review is on CVD and type 2 diabetes mellitus (T2DM), diseases with shared pathophysiology for which established risk assessment tools have influenced clinical practice.

EMERGENCE OF OMIC TECHNOLOGIES FOR CARDIOMETABOLIC DISEASE RISK PREDICTION

In 1961, Kannel et al. (1) reported on the first assessment of risk for coronary heart disease (CHD) in the Framingham Heart Study (FHS), an epidemiological cohort designed to investigate the risk factors for the development of CVD in the community. Clinical factors such as older age, male sex, and higher blood pressure and cholesterol were implicated in the development of disease (1). These clinical risk factors have been the mainstay of CVD risk assessment and are the central targets of preventative interventions in primary care and cardiology clinics (4). Risk assessment tools such as the one developed from the FHS have been useful in guiding clinical judgments and risk stratifying patients for optimal counseling and selection of therapy to improve disease risk profiles. Several tools have since been developed for CVD (5–7) and T2DM (8, 9), adding to the knowledge base of factors that influence the development of disease.

While a significant proportion of incident CMD is attributable to traditional risk factors, there remain individuals who develop disease without them (10). To optimize preventative therapy, there continues to be a need to identify novel markers and tools to improve risk prediction, and identify individuals with subclinical CMD and those at risk for future events. To improve clinical utility, candidate biomarkers should ideally have low correlation with established risk factors to improve risk prediction, as correlated biomarkers require the addition of significantly more analytes to obtain meaningful increases in model performance. As such, the broad-scale investigation of genes, proteins, and metabolites, among other omic measures, greatly enhances the ability to assess for orthogonal measures of disease risk, potentially adding diagnostic value to available risk assessments (11, 12). These new biological dimensions have the potential to significantly improve prediction

models for CMD while offering valuable insight into possible pathways and molecular changes associated with disease development.

ASSESSMENT OF CLINICAL UTILITY OF OMIC BIOMARKERS

Despite the enthusiasm surrounding omics for pathway discovery, translating findings from omic technologies to clinical practice has been difficult, and no single finding is sufficiently robust for clinical deployment. Central barriers to this process include standardization of assay measurements, external validation of key findings, and inclusion of diversified cohorts in study designs. Most importantly, candidate biomarkers need to demonstrate improvement of risk prediction over traditional risk factors. Several measures of prediction performance exist to evaluate candidate risk assessment tools. The most common metric, discrimination (area under the curve or C-Statistics), assesses the ability of the biomarker to correctly determine whether an individual gets the disease and incorporates the sensitivity and specificity of the test. Calibration (often measured by Hosmer-Lemeshow statistic) assesses the agreement between observed and predicted risk. Finally, the net reclassification index assesses the ability of the test to accurately move individuals across risk groups. This instrument is particularly useful to reclassify intermediate-risk patients, especially in regard to preventative therapies. While these measures are imperfect (13) and assessments across omic studies vary substantially, they begin to evaluate the potential utility of using candidate biomarkers in practice.

GENOMICS

The genetic contribution to the development of CMD was initially established by analysis of familial pedigrees and twin studies (14). It is generally estimated that for diseases such as CHD and T2DM, approximately 50% of disease risk is heritable; the other half is influenced by environmental factors. It is believed that the majority of cases of complex diseases such as CHD and T2DM are polygenic: Multiple genetic variants with small effect sizes contribute the development of disease in addition to environmental influences (15, 16). Genome-wide association studies (GWAS) have helped uncover novel risk loci, and the recent development of the polygenic risk score (PRS) has ushered in a new era of risk stratification for the genetic contribution to the development of CMD.

Single Risk Loci

For Mendelian or monogenic diseases, mutations in a single gene can explain phenotypic disease presentation. For instance, familial hypercholesterolemia (FH) can result from mutations in genes such as *LDLR* and *ApoB*. These mutations in turn result in substantial increases in the risk of CHD (>3-fold risk) (17). Similarly, monogenic diabetes mellitus or maturity-onset diabetes of the young (MODY) can result from single mutations in several genes including *HNF1A* (18). Incorporating these variants in risk assessments significantly improves prediction for complex diseases such as CHD and T2DM. Mendelian diseases are relatively few, however, and their contributions to the heritability estimates of CHD and T2DM are small. Much of the genetic contribution to CMD has been traced to variants with very small effect sizes. The remarkable completion of the Human Genome Project, and subsequent genotyping of epidemiological cohorts, led to the discovery of hundreds of novel loci associated with the development of CMD. For instance, single-nucleotide polymorphisms (SNPs) in chromosome location 9p21 have been consistently associated with increased risk for CHD (19). Identification of multiple genetic loci with small effect sizes that are associated with disease must be subject to stringent statistical significance testing (stringent *p*-values

for multiple hypotheses testing) to increase confidence in their predictive value. To increase the power for detecting genetic associations and to overcome stringent *p*-values for multiple hypotheses testing, large consortiums were established, pooling data from individual GWAS (20–22). The sharing and collaborative nature of these consortiums has proven instrumental in identifying hundreds of loci implicated in CMD. Still, the effect sizes for these genetic variants are small, and risk prediction for individual variants over traditional risk factors has been marginal. Further, the identified disease loci explain only a modest fraction of the heritable component of CMD, and some have hypothesized that the "missing heritability" lies in the potential discovery of rare variants with larger effect sizes (23).

Polygenic Risk Scores

The observation of small effect sizes for individual genetic variants has led investigators to aggregate common SNPs into disease risk scores. Early studies relating polygenic risk to CMD provided insight into high-risk groups with substantially larger risk for development of disease than individual SNP contributions would have predicted, i.e., a high PRS. Assessing 18 loci associated with T2DM, Lango et al. (24) identified a small group of individuals in the GoDARTS cohort with >24 risk alleles, conferring a greater-than-fourfold increase in risk for development of T2DM as compared to individuals with 10–12 risk alleles (**Table 1**). This aggregate risk was substantially higher than that conferred by individual SNPs. However, the information from all the genetic variants implicated in T2DM only marginally improved discrimination over a baseline model comprising readily available clinical factors including age, sex, and body mass index (24). Similar marginal improvements in risk prediction over clinical risk factors were seen for CHD in early risk scores of aggregate SNPs (25, 26).

Subsequent GWAS analyses of larger cohorts have greatly increased the number of known risk variants for CMD and improved risk prediction over standard clinical factors (21, 26–29). A PRS comprising 50 genetic variants associated with CHD identified a 91% increase in risk in the top quintile of subjects compared with the bottom quintile in a study of multiple large epidemiological cohorts (27). Extending the PRS to include variants below genome-wide significance, Khera et al. (30) analyzed 6.6 million SNPs and demonstrated that the individuals in the top 2.3% of PRS had a greater-than-fourfold increase in risk for CHD as compared to the rest of the population. While the improvements in risk prediction over standard risk factors have been modest and at times inconsistent, the inclusion of many more variants in a PRS, along with significant increases in sample size, has resulted in greater separation for risk groups at the tail ends of the population (30–32) (**Table 1**).

Limitations and Challenges for Use of Polygenic Risk Scores

The PRS is poised to be an important tool for predicting risk of CMD. However, knowledge of the limitations of PRSs is needed to avoid pitfalls associated with their potential use in clinical practice. One of the critical issues in genomic research is the general lack of diversity and representation of historically underrepresented groups in study designs. Individuals of European descent make up almost 80% of all GWAS subjects yet represent only 16% of the global population. Given differences in patterns of linkage disequilibrium and gene–gene and gene–environment interactions, prediction accuracies of a PRS are lower for underrepresented populations because of the makeup of the reference cohorts (33). Efforts are under way to increase the genetic diversity of GWAS cohorts to help alleviate this problem. The limitations associated with using GWAS-based methods for prediction of complex diseases such as CMD will hopefully be mitigated as whole-genome sequencing becomes more feasible and less cost-prohibitive across different populations.

Table 1 Scale of genetic variants used for polygenic risk scores and impact on cardiometabolic disease risk prediction

CMD	Cohorts	Risk score	Findings	Model discrimination	Reference
DM	GoDARTS: 2,309 cases/2,598 controls	SNPs in 18 loci at GWS	>24 risk (1.2%) alleles OR 4.2 (95% CI 2.11–8.56) versus 10–12 risk alleles (1.8%)	C index: 0.60 (PRS) versus 0.78 (RF) versus 0.80 (RF + PRS)	24
DM	Derivation: 32 studies, 74,124 cases/824,006 controls Validation: UK Biobank	136,795 SNPs	Top 2.5% of PRS had 3.4 times risk of median PRS	C index: 0.66 for RF versus 0.66 for PRS	32
CHD	Prospective cohort: 30,725 individuals from Sweden and Finland	13 SNPs	Top quintile versus bottom quintile: HR, 1.66 (95% CI 1.35–2.04)	C index: no improvement ($p = 0.19$) NRI: no improvement ($p = 0.18$)	26
CHD	Derivation: CARDIoGRAMplusC4D Consortium: 64,746 cases/130,681 controls Validation: FINRISK and FHS: 1,344 cases/16,082 controls	49,310 SNPs	Top quintile versus rest of population: OR 1.94 (95% CI 1.85–2.03)	C index: +1.6% increase with PRS (95% CI 0.01–0.02) over FRS	31
CHD	Derivation: established GWAS Validation: UK Biobank, 3,963 cases/120,280 controls Testing: 8,676 cases/288,978 controls	6.6 million SNPs	Top quintile versus rest of population: 2.55 (95% CI 2.43 versus 2.67)	C statistic: 0.81 in testing set; no comparison to RF	30

Abbreviations: CHD, coronary heart disease; CI, confidence interval; CMD, cardiometabolic disease; DM, diabetes mellitus; FHS: Framingham Heart Study; FRS, Framingham Risk Score; GWAS, genome-wide association study; GWS, genome-wide significance; HR, hazard ratio; NRI, net reclassification index; OR, odds ratio; PRS, polygenic risk score; RF, risk factor; SNP, single-nucleotide polymorphism.

There is concern that the use of the PRS in clinical practice might convey a deterministic attitude for at-risk individuals in regard to lifestyle modifications. However, it is important to note that the PRS has been shown to be adaptable to lifestyle choices, where high-risk individuals can alter their risk with healthy behaviors (27).

The excitement around using the PRS in clinical practice must be met with realistic expectations. The merits of the PRS are clear, as it is an unbiased tool to elucidate subgroups of individuals with heightened risk of disease. However, PRSs are not strong screening tools, as they poorly predict who actually develops disease or not (34). The same can be said for traditional clinical factors, and as such, the consideration of the score as a risk factor alongside comprehensive evaluation for prediction of disease may be the most responsible use for this emerging tool.

PROTEOMICS

Downstream of genes, proteins provide the functional readout of the genome and are the main effectors of cellular function. Further, genes that may harbor risk-carrying alleles may poorly predict disease, given influences from epigenetic mechanisms or environmental factors such as diet on

Table 2 Characteristics of proteomic methods in cardiometabolic biomarker discovery (scale: ✔ to ✔✔✔✔)

Proteomic method	Description	Sensitivity	Specificity	Proteome coverage	Throughput
Mass spectrometry	Cleavage, ionization of peptides, and measurement of mass/charge ratios	✔✔✔✔	✔✔✔✔	✔✔✔✔	✔
Multiplex-antibody	Bead- or microarray-based assays	✔✔✔✔	✔✔✔	✔	✔✔✔
Multiplex-antibody (PEA)	Antibody pairs with tag oligonucleotides and subsequent PCR	✔✔✔	✔✔✔	✔✔	✔✔✔
Multiplex-aptamer	Short oligonucleotides as affinity reagents in place of antibodies	✔✔✔	✔✔	✔✔✔	✔✔✔✔

Abbreviations: PEA, proximal extension assay; PCR, polymerase chain reaction.

downstream protein production. Proteomics, the large-scale study of proteins in the blood or tissues, is considerably more complex than genomics because of the dynamic nature of the synthesis, breakdown, and post-translational modifications of proteins (35). This has made the application of proteomics to both uncover novel mechanisms and assess CMD risk particularly challenging. Evolution of proteomic technologies in the recent past has sought to overcome the difficulties in capturing the breadth of proteomic discovery, particularly in the plasma where protein concentrations span over 10 logs, while preserving specificity for analysis of candidate biomarkers (35–37). Proteomics offers particular promise in providing a more comprehensive assessment for CMD risk prediction, integrating both genetic and environmental factors for disease development (36).

Mass Spectrometry

The goal of proteomics is to capture the breadth of proteins associated with disease processes (38). Methods such as mass spectrometry (MS) have been developed and refined to increase the amount of protein simultaneously measured from samples. After proteins are initially cleaved to peptides and separated via techniques such as liquid chromatography (LC), ionized peptides can be measured and quantified using their mass-to-charge ratios in MS-based techniques (39). Masses are then matched to libraries of peptides to deduce the protein of interest for precise identification of the analyte of interest. Throughput issues have limited the use of MS-based techniques in the study of large epidemiological cohorts (**Table 2**), as they have been best suited for studies of cellular biology, small perturbational experiments (40), and small-cohort designs (41).

Affinity-Based Reagents

The gold standard for protein identification involves immunoassay-based techniques, relying on affinity-based reagents such as antibodies for detection and quantification in biological samples. These methods include the enzyme-linked immunosorbent assay (ELISA), optimal for measurement of single protein analytes. Multiplex assays using affinity-based reagents (antibody or DNA aptamers) are capable of measuring multiple proteins at once in targeted proteomics and have increased throughput, allowing analysis of large epidemiological cohorts (35, 42–46). A multiplex protein panel of 85 proteins was used to ascertain associations between proteins and development of incident CVD in the 3,523 individuals in the FHS. Of the 8 proteins independently associated with incident CVD, growth differentiation factor-15 (GDF-15), a stress- and immune-responsive cytokine, remained significant in a multi-marker adjustment and added marginal discriminatory ability to the traditional risk factor model for CVD (47).

Despite the success of multiplexed antibody-based techniques in increasing throughput and simultaneous protein assays, these methods are limited in scope for proteomic analysis due to concerns of cross-reactivity and decreases in specificity. To overcome these issues, a proximal extension assay (PEA) technology has been developed that provides multiplex protein measurements with the attachment, hybridization, and subsequent polymerase chain reaction amplification (for quantification) of protein-specific oligonucleotide tags to antibody pairs. This provides more specific affinity reagents and allows simultaneous quantification of thousands of circulating proteins, increasing proteomic coverage for CMD biomarker discovery (43). Finally, a multiplexed protein assay technique with the use of DNA aptamers (single short-stranded oligonucleotides with high affinity for target proteins) as affinity reagents has expanded the scale of proteomic investigation for large epidemiological cohorts. In proof-of-principle studies, this approach has found novel protein biomarkers of acute myocardial injury, including several whose levels increase prior to established markers such as the cardiac troponins (48). This technology was applied to a cohort of patients with stable ischemic heart disease to predict cardiovascular events and compared with the Framingham secondary event model to determine predictive accuracy. Out of 1,130 proteins measured, a nine-protein risk model was derived, which, when added to the Framingham risk model, moderately improved discrimination performance (49). This risk score has been applied in a clinical trial, predicting harm from a lipid-lowering medication within three months of initiation of drug (50), highlighting the potential of multi-marker scores to improve risk prediction in a variety of settings.

While multiplex affinity-based techniques have enhanced potential for discovery, there is concern that the potential lack of analytic specificity may be a detriment to biomarker discovery. Thus, it is vital that candidate markers brought forward for CMD risk prediction be verified using standard MS-based techniques for precise measurements, while investigators simultaneously assess whether the new markers add information on top of what is known from existing clinical factors.

METABOLOMICS

Metabolites are small molecules (<1 kDa) or chemical intermediates of biological processes. Because they are downstream of genes and proteins, they represent proximal surrogates for observed clinical phenotypes. The broad-scale study of metabolites, or metabolomics, provides the ability to assess contributions from other omic fields along with environmental influences that contribute to disease processes (51). Metabolomic investigations attempt to cover the breadth of metabolic pathways, a challenging goal given the tens of thousands of metabolites in the metabolome. The two most widely adopted methods for study are nuclear magnetic resonance and LC/MS-based approaches (51, 52), with variations in techniques to optimize profiling of metabolite classes including sugars, amino acids, and lipids, among others. Similar to proteomics, metabolomics can follow a targeted approach in which compounds (known standards) are purchased and then assays are tuned on a mass spectrometer to develop multiplexed assays that leverage the unique mass and the retention time on a specific column. By contrast, newer sensitive instruments can more readily assay a broader swath of the large molecular diversity in a given complex mixture with an agnostic approach. Whereas targeted methods ensure unambiguous molecular identification at the outset, nontargeted methods promote serendipity. However, the identification of an MS peak of interest can prove arduous (**Table 3**). The scope of studies to date has been relatively modest when compared to genetic investigations, although throughput continues to improve. There is emerging consensus that these techniques are highlighting new pathways that may contribute to CMD pathogenesis that could also serve as disease biomarkers, highlighted by associations observed with amino acids and lipid metabolites.

Table 3 Characteristics of metabolomic methods in cardiometabolic biomarker discovery (scale: ✔ to ✔✔✔✔)

Metabolomic method	Description	Sensitivity	Specificity	Metabolome coverage	Throughput
NMR	Nondestructive; excitation of nuclei generates electronic signatures of sample compounds	✔	✔✔✔✔	✔	✔✔✔✔
Targeted LC/MS	Known compounds serve as internal standards with tuning of MS to unique mass and retention time of metabolite	✔✔✔✔	✔✔✔✔	✔✔	✔✔✔
Non-targeted LC/MS	Nonbiased profiling; necessitates peak identification	✔✔✔	✔✔✔	✔✔✔✔	✔✔✔

Abbreviations: NMR, nuclear magnetic resonance; LC, liquid chromatography; MS, mass spectrometry.

Amino Acids and Cardiometabolic Disease Risk

Elevated levels of branched-chain amino acids (BCAAs) have shown a consistent link to the development of CMD, and in particular, T2DM. In a case-control study, a metabolomic screen of amino acids and other polar metabolites found five amino acids associated with the development of T2DM in the FHS, more than a decade prior to onset of disease, while providing incremental improvement in model discrimination over standard diabetic risk factors. A score consisting of three metabolites (two BCAAs and one aromatic amino acid) was associated with a fivefold increase in the risk of T2DM in the top quartile compared to the bottom quartile. These findings were replicated in independent cohorts and provide a robust association for BCAAs in the pathogenesis of T2DM (53). Further, a potential causal association between impaired BCAA catabolism and the development of T2DM has been identified by leveraged metabolite and genetic information in so-called Mendelian randomization approaches (54). BCAAs have also been associated with the development of CHD, although insulin resistance and T2DM may partly mediate this association (55). Several other amino acids and derivatives including lysine, betaine, and glutamate have been implicated in CMD (56–59), highlighting the importance of this class of metabolites for risk prediction, and studies in heterogeneous populations are ongoing.

Lipids and Cardiometabolic Disease Risk

While dyslipidemia is a hallmark risk factor for CVD, an expanded investigation of perturbed lipid metabolism has revealed significant associations with the development of other diseases (60–64). New techniques capture information not obtainable by clinical tests, including distinct lipid classes or structural information including the length and saturation of lipids, which can be influenced by both dietary and genetic factors. A screen of 135 lipids revealed several associations between specific lipid species and incident CVD in the Bruneck Study. Specific cholesterol esters, phosphatidylcholines, and triacylglycerols were the strongest predictors of incident events and modestly improved risk discrimination over traditional risk factors, including total cholesterol and high-density lipoprotein cholesterol levels. As noted, lipids vary substantially, including in length and bond formations, and have been increasingly associated with both increased risk and protection against CMD based on species type. Metabolites of dietary lipids have also been implicated in the development of CHD. An untargeted metabolomic investigation in patients with high risk of developing CVD found associations of 18 metabolites with incident disease. Three correlated dietary metabolites of phosphatidylcholine, namely choline, trimethylamine N-oxide,

and betaine, were strongly associated with the development of disease. Experimental studies in mice found that suppression of intestinal microflora was associated with a decrease in choline-accelerated atherosclerosis, uncovering a possible link between gut microbiome metabolism and the development of CVD. Lipid species and their metabolites are implicated in a host of biological processes and may be mediators of residual risk in the setting of controlled clinically measured cholesterol levels.

Application of metabolomic screening at a more targeted level has been instrumental in assessing for various clinical scenarios, including inborn errors of metabolism (65). Metabolomic technologies have improved the ability to identify and elucidate subtleties in phenotypes for these rare diseases. It is conceivable with improvements in technologies that these methods can be applied to correlate thousands of metabolites to broad ranges of cardiometabolic phenotypes to improve our ability to assess risk and refine risk prediction.

MULTI-OMICS

While genomics, proteomics, and metabolomics form the core of omics research, several emerging omic fields are poised to further contribute to our understanding of CMD risk. Disciplines such as transcriptomics have helped highlight the function of RNA transcripts, and epigenomics has identified modifications to the genome including DNA methylation and histone modifications that contribute to disease development (66, 67). Further, while omics research has primarily analyzed blood for analytes of interest, significant advances have been made in the analysis of tissue samples, including the heart, liver, and gut microbiome (60), that may provide new insights into organ-specific biological processes involved in CMD pathology. Given the complementarity of information provided by the aforementioned areas of omic profiling, there is increasing enthusiasm for multi-marker panels or risk scores for disease prediction. Each omic field can provide novel prognostic information of disease risk. However, no single data set can explain the entirety of risk prediction as these fields provide orthogonal information. A multi-omic approach to CMD research would provide incremental improvements in risk prediction compared to standard clinical risk factors (68) while helping to unravel pathways or mechanisms of disease from genetic risk loci to phenotypes, thereby providing avenues to discover important biomarkers and therapeutic targets for future research (69, 70). Integration of these data sets with traditional risk factors holds the true potential for omics in uncovering powerful diagnostic tools for CMD research. However, given the complexity and high dimensionality of the data sets involved, traditional statistical approaches to data analysis for prediction may be insufficient. Advanced statistical modeling, including the use of machine learning, may hold promise in handling such complex data, as these methods have shown great potential in accounting for complex interactions across data sets and have helped improve risk prediction compared to traditional models (71). However, significant challenges lie ahead as omic data sets are heterogeneous, noisy, and often too small for effective implementation of these methods. Further, as the complexity of analysis continues to grow in order to deal with the vastness of omic data, translating findings that are interpretable for clinicians will present significant challenges for implementation to clinical practice.

CONCLUSION

The emergence of omic technologies has highlighted orthogonal pathways for CMD development, producing a plethora of novel potential biomarkers and tools for risk assessment. However, the path toward clinical utility for these measures is arduous. Several key questions need to be addressed prior to bringing candidate risk tools or biomarkers forward. Goals include assessment

of the clinical utility that biomarkers can add to traditional risk factors, validation of findings in diverse cohorts, and standardization of assay technologies, among others. Nonetheless, the recent past has seen extraordinary developments in the broad-scale and systematic investigation of novel risk factors for disease. Only two decades ago, it was inconceivable that we would be sequencing human genomes and integrating genomic information into clinical practice. There is optimism that these new profiling technologies will play a role in the clinical assessment of CMD risk.

FUTURE ISSUES

1. Validation of candidate omic analytes or genetic risk tools in relation to disease in external cohorts. This includes addressing disparities in cohort designs to clarify differentiation of risk across sexes and races.
2. Comprehensive analysis for utility of candidate biomarkers or risk prediction tools to improve risk prediction over traditional risk factors for CMD.
3. Standardization of omic methodologies. This presents a unique challenge, especially in relation to consistency of sample preparation and assay development. Pipelines for biomarker discovery and subsequent validation will be crucial to translate findings to clinical utility and practice (40, 72).
4. Recognition of practical considerations for biomarkers when assessing bringing candidate analytes forward to practice. These include (*a*) absolute quantification of biomarkers, as omic studies often standardize measurements to reference values providing relative concentrations, and (*b*) assessments for optimal cut points in defining abnormal ranges that indicate disease development.
5. Development and refinement of novel statistical methods to analyze and integrate high-dimensional omic data.
6. Assessment of the utility of omic biomarkers to improve risk prediction on the basis of response to therapy.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors acknowledge grant support from the National Institutes of Health under grant numbers R01DK081572, R01HL132320, and R01HL133870. We apologize in advance to all the investigators whose work could not be cited due to space limitations.

LITERATURE CITED

1. Kannel WB, Dawber TR, Kagan A, et al. 1961. Factors of risk in the development of coronary heart disease—six year follow-up experience. The Framingham Study. *Ann. Intern. Med.* 55:33–50
2. Morrow DA, de Lemos JA. 2007. Benchmarks for the assessment of novel cardiovascular biomarkers. *Circulation* 115:949–52

3. de Lemos JA, Rohatgi A, Ayers CR. 2015. Applying a big data approach to biomarker discovery: running before we walk? *Circulation* 132:2289–92
4. Wilson PW, D'Agostino RB, Levy D, et al. 1998. Prediction of coronary heart disease using risk factor categories. *Circulation* 97:1837–47
5. Goff David C, Lloyd-Jones DM, Bennett G, et al. 2014. 2013 ACC/AHA guideline on the assessment of cardiovascular risk. *Circulation* 129:S49–73
6. Ridker PM, Buring JE, Rifai N, Cook NR. 2007. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: The Reynolds Risk Score. *JAMA* 297:611–19
7. Hippisley-Cox J, Coupland C, Brindle P. 2017. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ Clin. Res. Ed.* 357:j2099
8. Bang H, Edwards AM, Bomback AS, et al. 2009. Development and validation of a patient self-assessment score for diabetes risk. *Ann. Intern. Med.* 151:775–83
9. Heikes KE, Eddy DM, Arondekar B, Schlessinger L. 2008. Diabetes Risk Calculator: a simple tool for detecting undiagnosed diabetes and pre-diabetes. *Diabetes Care* 31:1040–45
10. Khot UN, Khot MB, Bajzer CT, et al. 2003. Prevalence of conventional risk factors in patients with coronary heart disease. *JAMA* 290:898–904
11. Wang TJ. 2011. Assessing the role of circulating, genetic, and imaging biomarkers in cardiovascular risk prediction. *Circulation* 123:551–65
12. Hoefer IE, Steffens S, Ala-Korpela M, et al. 2015. Novel methodologies for biomarker discovery in atherosclerosis. *Eur. Heart J.* 36:2635–42
13. Pepe MS, Fan J, Feng Z, et al. 2015. The net reclassification index (NRI): a misleading measure of prediction improvement even with independent test data sets. *Stat. Biosci.* 7:282–95
14. Marenberg ME, Risch N, Berkman LF, et al. 1994. Genetic susceptibility to death from coronary heart disease in a study of twins. *N. Engl. J. Med.* 330:1041–46
15. Musunuru K, Kathiresan S. 2019. Genetics of common, complex coronary artery disease. *Cell* 177:132–45
16. Kathiresan S, Srivastava D. 2012. Genetics of human cardiovascular disease. *Cell* 148:1242–57
17. Abul-Husn NS, Manickam K, Jones LK, et al. 2016. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science* 354(6319):aaf7000
18. Estrada K, Aukrust I, Bjorkhaug L, et al. 2014. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *JAMA* 311:2305–14
19. Talmud PJ, Cooper JA, Palmen J, et al. 2008. Chromosome 9p21.3 coronary heart disease locus genotype and prospective risk of CHD in healthy middle-aged men. *Clin. Chem.* 54:467–74
20. Deloukas P, Kanoni S, Willenborg C, et al. 2013. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* 45:25–33
21. Nikpay M, Goel A, Won HH, et al. 2015. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* 47:1121–30
22. Liu C, Kraja AT, Smith JA, et al. 2016. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat. Genet.* 48:1162–70
23. Manolio TA, Collins FS, Cox NJ, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–53
24. Lango H, Palmer CN, Morris AD, et al. 2008. Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. *Diabetes* 57:3129–35
25. Morris RW, Cooper JA, Shah T, et al. 2016. Marginal role for 53 common genetic variants in cardiovascular disease prediction. *Heart* 102:1640–47
26. Ripatti S, Tikkanen E, Orho-Melander M, et al. 2010. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* 376:1393–400
27. Khera AV, Emdin CA, Drake I, et al. 2016. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N. Engl. J. Med.* 375:2349–58
28. Inouye M, Abraham G, Nelson CP, et al. 2018. Genomic risk prediction of coronary artery disease in 480,000 adults. *Implications Primary Prev*. 72:1883–93

29. Bonifacio E, Beyerlein A, Hippich M, et al. 2018. Genetic scores to stratify risk of developing multiple islet autoantibodies and type 1 diabetes: a prospective study in children. *PLOS Med.* 15:e1002548
30. Khera AV, Chaffin M, Aragam KG, et al. 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* 50:1219–24
31. Abraham G, Havulinna AS, Bhalala OG, et al. 2016. Genomic prediction of coronary heart disease. *Eur. Heart J.* 37:3267–78
32. Mahajan A, Taliun D, Thurner M, et al. 2018. Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* 50:1505–13
33. Martin AR, Kanai M, Kamatani Y, et al. 2019. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51:584–91
34. Wald NJ, Old R. 2019. The illusion of polygenic disease risk prediction. *Genet. Med.* 21:1705–7
35. Lam MPY, Ping P, Murphy E. 2016. Proteomics research in cardiovascular medicine and biomarker discovery. *J. Am. Coll. Cardiol.* 68:2819–30
36. Lindsey Merry L, Mayr M, Gomes Aldrin V, et al. 2015. Transformative impact of proteomics on cardiovascular health and disease. *Circulation* 132:852–72
37. Benson MD, Sabatine MS, Gerszten RE. 2016. The prospects for cardiovascular proteomics: a glass approaching half full. *JAMA Cardiol.* 1:245–46
38. Edwards AVG, White MY, Cordwell SJ. 2008. The role of proteomics in clinical cardiovascular biomarker discovery. *Mol. Cell. Proteom.* 7:1824–37
39. Han X, Aslanian A, Yates JR 3rd. 2008. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* 12:483–90
40. Addona TA, Shi X, Keshishian H, et al. 2011. A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nat. Biotechnol.* 29:635–43
41. Yin X, Subramanian S, Hwang SJ, et al. 2014. Protein biomarkers of new-onset cardiovascular disease: prospective study from the systems approach to biomarker research in cardiovascular disease initiative. *Arterioscler. Thromb. Vasc. Biol.* 34:939–45
42. Gerstein HC, Pare G, McQueen MJ, et al. 2015. Identifying novel biomarkers for cardiovascular events or death in people with dysglycemia. *Circulation* 132:2297–304
43. Lind L, Siegbahn A, Lindahl B, et al. 2015. Discovery of new risk markers for ischemic stroke using a novel targeted proteomics chip. *Stroke* 46:3340–47
44. Nowak C, Sundstrom J, Gustafsson S, et al. 2016. Protein biomarkers for insulin resistance and type 2 diabetes risk in two large community cohorts. *Diabetes* 65:276–84
45. Nowak C, Carlsson AC, Östgren CJ, et al. 2018. Multiplex proteomics for prediction of major cardiovascular events in type 2 diabetes. *Diabetologia* 61:1748–57
46. Molvin J, Pareek M, Jujic A, et al. 2019. Using a targeted proteomics chip to explore pathophysiological pathways for incident diabetes—the Malmo Preventive Project. *Sci. Rep.* 9:272
47. Ho JE, Lyass A, Courchesne P, et al. 2018. Protein biomarkers of cardiovascular disease and mortality in the community. *J. Am. Heart Assoc.* 7:e008108
48. Ngo D, Sinha S, Shen D, et al. 2016. Aptamer-based proteomic profiling reveals novel candidate biomarkers and pathways in cardiovascular disease. *Circulation* 134:270–85
49. Ganz P, Heidecker B, Hveem K, et al. 2016. Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. *JAMA* 315:2532–41
50. Williams SA, Murthy AC, DeLisle RK, et al. 2018. Improving assessment of drug safety through proteomics. *Circulation* 137:999–1010
51. McGarrah RW, Crown SB, Zhang G-F, et al. 2018. Cardiovascular metabolomics. *Circ. Res.* 122:1238–58
52. Rhee EP, Gerszten RE. 2012. Metabolomics and cardiovascular biomarker discovery. *Clin. Chem.* 58:139–47
53. Wang TJ, Larson MG, Vasan RS, et al. 2011. Metabolite profiles and the risk of developing diabetes. *Nat. Med.* 17:448
54. Lotta LA, Scott RA, Sharp SJ, et al. 2016. Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a Mendelian randomisation analysis. *PLOS Med.* 13:e1002179

55. Tobias DK, Lawler PR, Harada PH, et al. 2018. Circulating branched-chain amino acids and incident cardiovascular disease in a prospective cohort of US women. *Circ. Genom. Precision Med.* 11:e002157
56. Walford GA, Ma Y, Clish C, et al. 2016. Metabolite profiles of diabetes incidence and intervention response in the diabetes prevention program. *Diabetes* 65:1424–33
57. Vaarhorst AA, Verhoeven A, Weller CM, et al. 2014. A metabolomic profile is associated with the risk of incident coronary heart disease. *Am. Heart J.* 168:45–52.e7
58. Wurtz P, Havulinna AS, Soininen P, et al. 2015. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 131:774–85
59. Ruiz-Canela M, Hruby A, Clish CB, et al. 2017. Comprehensive metabolomic profiling and incident cardiovascular disease: a systematic review. *J. Am. Heart Assoc.* 6:e005705
60. Wang Z, Klipfell E, Bennett BJ, et al. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472:57–63
61. Ganna A, Salihovic S, Sundstrom J, et al. 2014. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLOS Genet.* 10:e1004801
62. Stegemann C, Pechlaner R, Willeit P, et al. 2014. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. *Circulation* 129:1821–31
63. Lawler PR, Akinkuolie AO, Chu AY, et al. 2017. Atherogenic lipoprotein determinants of cardiovascular disease and residual risk among individuals with low low-density lipoprotein cholesterol. *J. Am. Heart Assoc.* 6:e005549
64. Paynter NP, Balasubramanian R, Giulianini F, et al. 2018. Metabolic predictors of incident coronary heart disease in women. *Circulation* 137:841–53
65. Tebani A, Abily-Donval L, Afonso C, et al. 2016. Clinical metabolomics: the new metabolic window for inborn errors of metabolism investigations in the post-genomic era. *Int. J. Mol. Sci.* 17:1167
66. Ordovás JM, Smith CE. 2010. Epigenetics and cardiovascular disease. *Nat. Rev. Cardiol.* 7:510–19
67. Lu D, Thum T. 2019. RNA-based diagnostic and therapeutic strategies for cardiovascular disease. *Nat. Rev. Cardiol.* **https://doi.org/10.1038/s41569-019-0218-x**
68. Walford GA, Porneala BC, Dauriz M, et al. 2014. Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes. *Diabetes Care* 37:2508–14
69. Kullo IJ, Cooper LT. 2010. Early identification of cardiovascular risk using genomics and proteomics. *Nat. Rev. Cardiol.* 7:309
70. Shah SH, Newgard CB. 2015. Integrated metabolomics and genomics: systems approaches to biomarkers and mechanisms of cardiovascular disease. *Circ. Cardiovasc. Genet.* 8:410–19
71. Kakadiaris IA, Vrigkas M, Yen AA, et al. 2018. Machine learning outperforms ACC/AHA CVD risk calculator in MESA. *J. Am. Heart Assoc.* 7:e009476
72. Rifai N, Gillette MA, Carr SA. 2006. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* 24:971–83