91308_Auto_Edited-check.docx

WORD COUNT	3078	TIME SUBMITTED	21-MAR-2024 01:52PM
		PAPER ID	107720598

Name of Journal: *World Journal of Gastroenterology*

Manuscript NO: 91308

Manuscript Type: MINIREVIEWS

Omics-based biomarkers as useful tools in metabolic dysfunction-associated steatotic liver disease clinical practice: How far are we?

Trinks J *et al.* Omics-based biomarkers in MASLD

Abstract

Unmet needs exist in metabolic dysfunction-associated steatotic liver disease (MASLD) risk stratification. Our ability to identify patients with MASLD with advanced fibrosis and at higher risk for adverse outcomes is still limited. Incorporating novel biomarkers could represent a meaningful improvement to current risk predictors. With this aim, omics technologies have revolutionized the process of MASLD biomarker discovery over the past decades. While the research in this field is thriving, much of the publication has been haphazard, often using single-omics data and specimen sets of convenience, with many identified candidate biomarkers but lacking clinical validation and utility. If we incorporate these biomarkers to direct patients' management, it should be considered that the roadmap for translating a newly discovered omics-based signature to an actual, analytically valid test useful in MASLD clinical practice is rigorous and, therefore, not easily accomplished. This article presents an overview of this area's current state, the conceivable opportunities and challenges of omics-based laboratory diagnostics, and a roadmap for improving MASLD biomarker research.

Key words: Metabolic dysfunction-associated steatotic liver disease; Non-alcoholic steatohepatitis; Biomarker; Risk stratification; Omics

Trinks J, Mascardi MF, Gadano A, Marciano S. Omics-based biomarkers as useful tools in metabolic dysfunction-associated steatotic liver disease clinical practice: How far are we? *World J Gastroenterol* 2024; In press

Core tip: Identifying patients with metabolic dysfunction-associated steatotic liver disease (MASLD) at higher risk for adverse outcomes is still a crucial clinical challenge. Novel and non-invasive screening, monitoring, and risk stratification methods are urgently needed. With this aim, omics technologies have revolutionized the process of MASLD biomarker discovery. Although many omics-based biomarkers were identified over the past decades, their translation into clinically useful tests that can guide management decisions has proven more difficult than expected. This review

presents an overview of this area's current state, the conceivable opportunities and challenges of omics-based laboratory diagnostics, and a roadmap for improving MASLD biomarker research.

INTRODUCTION

The exponential global increase in obesity and type 2 diabetes is to blame for the rising epidemic of metabolic dysfunction-associated steatotic liver disease (MASLD) in the last two decades. Globally, nearly a quarter of the world's population is estimated to be affected by MASLD, with even higher rates in the Middle East, Northern Africa, and Central and South America. Without clear risk stratification and therapeutic options, MASLD has soon become the leading cause of chronic liver disease and liver-related morbidity and mortality worldwide[1].

Unfortunately, over half the adult population is expected to have MASLD by 2040, mainly affecting women, smokers, and those without metabolic syndrome[2]. This scenario will expand the number of patients with advanced liver fibrosis and end-stage liver disease, increasing the rates of liver transplantation and multiplying healthcare costs.

Current measures to overcome this situation are centered on the identification of people at the highest risk of progression to advanced liver disease so they can be offered interventions and appropriate care before they develop liver-related complications[3]. However, this represents a crucial and tough challenge in the clinical management of MASLD[4-6]. First, not all patients with MASLD will ever develop non-alcoholic steatohepatitis (NASH) with clinically significant fibrosis, and advanced liver fibrosis has the highest risk of adverse liver-related outcomes[7]. Second, liver biopsy is the best available method to evaluate liver fibrosis. Still, it is a costly and invasive technique, prone to sampling bias and intra-observer and inter-observer variability, and not recommended to screen liver disease progression[8,9].

These limitations have driven the need for non-invasive MASLD screening, monitoring, and risk stratification methods. In this regard, imaging methods such as ultrasound, computed tomography, and magnetic resonance are valuable for detecting

lipid accumulation in MAFLD patients. Still, they are useless for evaluating inflammation and degrees of fibrosis less than cirrhosis[10,11].

On the contrary, low-tech methods such as the anthropometric clinical indicators of visceral obesity have been promising for the primary prevention and screening of MASLD. These non-invasive indicators, such as body mass index, abdomen, waist and chest circumferences, and trunk fat, can be easily determined to screen the presence or absence of MASLD using affordable equipment, which makes them ideal for their use in remote areas or in primary clinical practice[12,13]. Although these indicators show suboptimal accuracy (57%) in detecting fibrosis[13], the identification of their specific cutoff points, taking into consideration the anthropometric differences in each ethnic and racial group[12], may be useful in the early diagnosis of MASLD, and especially NASH, offering a new therapeutic and preventive method to this population. Therefore, the epicenter of all research efforts around MASLD is currently focused on identifying and validating new non-invasive and cost-effective biological markers (or biomarkers) for risk stratification[4].

Since the Human Genome Project, the development of new technologies called "omics" has made it possible to measure a vast number of biological molecules within a tissue or cell in a high-throughput way. Many areas of research can be classified as omics, and the terms used to define them depend on the type of biological molecules globally analyzed by these techniques; for example, genes (genomics), methylated DNA, or modified histone proteins in chromosomes (epigenomics), RNA (transcriptomics), proteins (proteomics), metabolites (metabolomics), lipids (lipidomics)[14]. These techniques require next-generation sequencing (NGS) or mass spectrometry (MS). While sequencing-based approaches are applied to study the genome, transcriptome, and epitomes, an MS is necessary to explore the proteome, metabolome, and lipidome[14].

In MASLD research, new innovative omics technologies are extensively used in both preclinical (*in vitro* and *in vivo*) models and retrospective studies with archived samples, as they offer the possibility of in-depth screening for novel biomarkers and a better understanding of MASLD pathological processes[15]. However, translating their results into clinically useful tests that can guide management decisions has proven

more difficult than expected. This article presents an overview of this area's current state, the conceivable opportunities and challenges of omics-based laboratory diagnostics, and a roadmap for improving MASLD biomarker research.

CURRENT STATE OF OMICS-DERIVED BIOMARKER DEVELOPMENT IN MASLD

We used PubMed (National Library of Medicine) and *Reference Citation Analysis* (*RCA*) databases to search and retrieve scientific articles to describe the current state of omics-derived biomarker development in MASLD. The "OR" and "AND" connectors were used to combine the descriptors: ("non-alcoholic fatty liver disease" or "non-alcoholic steatohepatitis", "fatty liver" or "NAFLD" or "NASH" or "metabolic-associated fatty liver disease" or "MAFLD" "metabolic dysfunction–associated steatotic liver disease" or "MASLD") and ("human") and ("biomarker") and ("omics" or "multi-omics" or "microbiome" or "genomics" or "proteomics" or "metabolomics" or "metagenomics" or "transcriptomics").

Inclusion criteria were the availability of the full-text publication written in English up to December 2023. Review studies, publications present in more than one database, and out-of-scope studies were excluded. After reading the titles and abstracts of the studies, 24 of 163 studies found in the databases were excluded: 2 duplicate studies, 5 Chinese studies, 11 *in vitro* studies, and six evaluated individuals with other chronic liver diseases. The remaining studies were fully read according to the eligibility criteria. Thus, 139 articles were included in the analysis. The complete list of articles included in the analysis and their primary information (authorship and year of publication, type of sample, and omic technique/s used in the study) are shown in Supplementary Table 1.

Primarily, omic papers on MASLD biomarker research were scarce 20 years ago. Since then, the number of omic articles has increased exponentially (Figure 1A). This systematic literature review showed that the recent and scanty publication of multi-omic studies focuses on three levels of omics data including transcripts, genes, and proteins. This is followed by other omics areas, comprising metabolites, epigenetic changes, and combinations thereof (Figure 1B).

Of the types of samples analyzed for MASLD biomarker research, the rise in the number of omic studies was followed by a boost of papers aimed at discovering MASLD biomarkers on a diverse array of biological samples (Figure 1C). However, liver tissue samples were the most frequently featured. Stool, plasma, and serum samples have dominated over the last few years (Figure 1C). These results indicate a current interest in the search for non-invasive biomarkers, primarily focusing on gut microbiota studies and its well-described relationship to MASLD development and progression.

The 139 studies were also analyzed according to a previously published set of criteria that evaluate the internal validity of individual studies[16,17]. The most common limitations observed in these studies were a small sample size (16.5%), followed by the presence of bias in the selection and stratification of patients due to the lack of histological diagnosis of MASLD using liver biopsies (12.2%). Moreover, disregard for potential confounding variables (such as age, gender, ethnicity, body mass index, or presence of comorbidities) and their appropriate adjustments in the data analysis was observed in 13 studies (9.35%).

The validity of a study can also be evaluated by its reproducibility. In the case of omics research, depositing raw data, complete protocols, and bioinformatics codes and workflows in a public repository is considered a first step to replicating a study's findings[17]. Although many leading journals now demand to make data and protocols publicly available as a prerequisite for publication[18], this practice remains inconsistent across journals and omics studies. In fact, seven (5%) of the studies included in our literature search did not offer public access to their data nor indicate how others may obtain it in case specific legal or ethical restrictions prohibit public sharing of the data set.

However, these flaws aren't recurrently detected in some fields of omics research. In the case of genomics, genome-wide association studies require high degree of certainty in the results (P values $< 5 \times 10^{-8}$), develop large, multi-centric replication studies, and have good compliance with data availability policies[19], as was observed in the two studies of this type (1.4%) analyzed in the literature search carried out in this review.

ADDRESSING THE LACK OF CLINICALLY USEFUL OMICS-BASED BIOMARKERS IN MASLD: WHAT ARE THE CHALLENGES?

Despite this progress, the omics-derived biomarker research in MASLD has resulted in several candidate biomarkers that lack clinical validation and utility; that is, it is yet unknown if the identified biomarkers are accurate, reproducible, or reliable in terms of the analytical and clinical/biological validation, or even if there is enough evidence to consistently demonstrate that the use of the omics-based predictor results in a better outcome for patient care (utility).

There are many reasons for this disappointing output. First, the advent of omics-based biomarker studies has led to an excess of highly confounded reports by a superlative number of variables applied to a small sample size. Conditions for sample collection, transport, and storage can significantly affect the quantity and quality of the molecules (nucleic acid, proteins, metabolites) to be analyzed in the biological sample[20]. Moreover, using different NGS platforms or reagents with dissimilar lot numbers can be a source of technical bias that can alter results[21].

Second, the analysis of an outstanding number of data can be challenging. This is especially true if no consensus exists on the data processing pipeline or the software and packages to be used[22,23]. In the case of those omics studies focused on the search for MASLD biomarkers in the human microbiome, the need for complete and curated databases for the human microbiota's less-studied viral and fungal components poses an obstacle to thorough data analysis[24].

Finally, the shift from a single-omic to a multi-omic approach to biomarker research is critical for increasing the chances of identifying an accurate biomarker for MASLD risk stratification. The number of multi-omics studies on MASLD biomarker research is still scarce but on the rise, as shown by the systematic literature review results. Multi-omic applications provide novel insights and a more holistic understanding of biological processes. Thus, this type of omic study could advance our ability to understand and treat the complex underlying biology of MASLD[25]. Unfortunately, the need for a rigorous study design, accurate sample size and statistical power

calculation, and problematic data integration interfere with the expansion of these studies[26].

A ROADMAP FOR THE IMPROVEMENT OF BIOMARKER RESEARCH IN MASLD

A lesson can be learned from cancer research. As a consequence of the premature use of omics-based tests developed to predict sensitivity to chemotherapeutic agents in lung and breast cancer clinical trials at Duke University, a committee of the Institute of Medicine generated a series of recommendations that are considered a roadmap of the best scientific practices for the development, validation and clinical translation of omics-based biomarkers and tests[27]. In 2013, the United States National Cancer Institute proposed a practical guideline[28,29] that lists 30 criteria for the development path of omics-based predictors from high-throughput technology to clinical trials (Figure 2). In brief, researchers should take into account: (1) Clinical specimen issues: Collection, processing, storage conditions, availability, quality, amount (mass or volume), and composition of appropriate clinical specimens; (2) Assay issues: Standardization of technical protocols, reagents, and scoring and reporting methods required for the analytical performance of the omics assay in the clinical setting; (3) Model development and evaluation: Avoidance of errors, inconsistencies, or bias in approachs for omics data pre-processing, preparation of the mathematical predictor model, and evaluation of its performance (validation); (4) Clinical trial design: Adherence to accepted standards for good clinical practice, including the development of a formal protocol, an informatics workflow for the analysis of clinical and omics data, a pre-defined study plan and statistical analysis strategy, and the pre-registration of the study and analysis plan in a public registry. The three possible study designs to evaluate the clinical utility are: Prospective-retrospective studies using stored samples; prospective clinical studies, where the biomarker is not involved in patient management decisions; or prospective clinical studies, where the biomarker guides patient management decisions; and (5) Ethical, legal, and regulatory issues: Commitment to protecting human subjects involved in the research, performance of certified laboratory tests if the results have significant clinical value and need to be

communicated to the patient or the patient's physician, documentation of intellectual property rights.

Although there are subtle differences in how this checklist is applied to a particular omics test and clinical setting, it is recognized that any field of omics-based biomarker research should take notice of the abovementioned criteria to determine when there is sufficient and reliable evidence to justify the clinical use of an omics-based biomarker, or even that it is ready for evaluation in a clinical trial.

After demonstrating its clinical usefulness, the new biomarker must prove its cost-effectiveness or cost-utility, mainly as its use will be widespread (up to 30% of the general population in some geographical regions). The benefit the application of the biomarker produces, measured by healthy life years' indicator, must exceed the cost of the intervention. The application of the biomarker is considered cost-effective if it produces at least USD 50000 per quality-adjusted life year gained[30].

WILL SERUM BIOMARKERS WIN THE RACE FOR CLINICAL USEFULNESS?

Regarding MASLD biomarker research, a significant step forward was recently taken by the non-invasive biomarkers for metabolic liver disease (NIMBLE) and liver investigation: testing marker utility in steatohepatitis (LITMUS) projects, as multiple circulating biomarkers made the first step in the biomarker qualification path[31,32].

The LITMUS project[32] evaluated the diagnostic accuracy of 5 single biomarkers (CK-18 M30, CK-18 M65, PRO-C3, PRO-C4, and PRO-C6), 9 multimarker scores (FIB-4, MACK-3, a scoring system proposed by Cao and colleagues in 2013, ADAPT, FIBC3, ABC3D, NFS, ELF, and SomaSignal), as well as liver stiffness measurement (LSM) and controlled attenuation parameter vibration-controlled transient elastography (VCTE) in detecting at-risk NASH and fibrosis severity in a European cohort of 966 biopsy-proven participants with MASLD (of which 335 patients had at-risk NASH and 271 advanced fibrosis).

None of the single markers or multimarker scores achieved the predefined acceptable area under the curve (AUC) of 0.8 to be considered a diagnostic marker of acceptable accuracy and for replacing biopsy in detecting people with both NASH and clinically significant fibrosis. However, the SomaSignal test, the ADAPT score, and the

LSM VCTE could be used as prescreening methods in clinical trials. The SomaSignal test showed the best results, with AUC values higher for diagnosing advanced fibrosis than for detecting NASH and clinically significant fibrosis[32].

On the other hand, the NIMBLE project tested the performance metrics of 5 biomarker panels (NIS4, OWLiver, PROC3, ELF, and FibroMeter VCTE) for the diagnosis of NASH, at-risk NASH or fibrosis severity in an American cohort of 1073 biopsy-proven individuals with the full spectrum of MASLD[31].

In this case, panels with an area under the receiver operating characteristic of 0.7 or higher were considered a diagnostic marker of acceptable accuracy. In contrast, primacy over alanine aminotransferase (ALT) and fibrosis-4 (FIB-4) for disease activity and fibrosis severity were considered, respectively, as a pragmatic initial step to move to final qualification. The results demonstrate that the NIS4 score excedeed the prespecified performance metric for diagnosis of at-risk NASH. Also, in diagnosing clinically significant fibrosis, advanced fibrosis, and cirrhosis in individuals with MASLD, the ELF test and FibroMeter VCTE outperformed FIB-4[31].

However, despite several limitations (data not applicable to all ethnicities, the use of a curated patient population, limited quantity of sample material, lack of evaluation of omics-based biomarkers, *etc.*) added to the known inter-observer variability in the histology scoring of the reference method, these groundbreaking studies represent an advance towards having regulatory approved biomarkers for MASLD risk stratification[5,6]. Moreover, as many tested biomarkers outperformed laboratory tools routinely used in the clinical practice, such as ALT and FIB-4, they could be applied in a pre-screening strategy in clinical trial recruitment[5,6].

CONCLUSION

The expectation of omics technologies is that in the future, patients might be diagnosed and treated according to their personalized MASLD molecular signatures. However, we have a long and arduous path to reach our goal. A consensus regarding best practices on sampling conditions, technical issues, and data processing on discovery studies is mandatory before omics-based biomarkers for MASLD can be validated and their clinical utility tested.

The increasing burden of MASLD emphasizes the pressing need for a novel biomarker that can surpass the "imperfect" liver histology. Moreover, it should be simple, cheap, and easy to adapt to different situations (population screening or stratification in clinical trials). Due to the complex multisystemic pathophysiology of MASLD, it seems unlikely that a single biomarker featuring all these attributes will be identified in the near future. On the contrary, different algorithms integrating clinical data with an arrangement of previously reported omics-based, circulating, anthropometric, and/or imaging-based markers could be considered strong candidates for clinical evaluation.

91308_Auto_Edited-check.docx

ORIGINALITY REPORT

2%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.bioclinica.com Internet	18 words — 1%
2	boris.unibe.ch Internet	16 words — < 1%
3	Lisa M McShane, Margaret M Cavenagh, Tracy G Lively, David A Eberhard et al. "Criteria for the use of omics-based predictors in clinical trials: explanation and elaboration", BMC Medicine, 2013 Crossref	15 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 15 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 15 WORDS