Universiteit
Leiden
The Netherlands

Multi-omics in research: epidemiology, methodology, and advanced data analysis

Faquih, T.O.

Citation
Faquih, T. O. (2023, March 28). *Multi-omics in research: epidemiology, methodology, and advanced data analysis*. Retrieved from https://hdl.handle.net/1887/3589838

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3589838

Note: To cite this publication please use the final published version (if applicable).

Chapter 8
Discussion and future perspectives

1 AIMS OF THIS THESIS

In this thesis, we explored epidemiological applications and methodological challenges of genomics, proteomics, and metabolomics. Metabolomics, one of the more recent OMICs fields, was the main focus of our research. Metabolites are thought to reflect integrated genomic and proteomic influences in metabolism as well as the environmental and external effects from the individuals' exposures. Hence, metabolomics may provide novel insight in the pathophysiology of complex multifactorial diseases. Indeed, in our research, metabolomics shed light on the associations between metabolites and non-alcoholic fatty liver disease (NAFLD), metabolites and short sequence repeats in the huntingtin gene, and the associations of per-/polyfluoroalkyl substances (PFAS) chemicals in the general population with metabolites. However, methodological, technological, and statistical challenges remain in this infant field. Therefore, we explored the issues of handling missing values in metabolomic data using a simulation study based on real data. This resulted in a publicly available R script to streamline the imputation of missing values of metabolites. In addition, we showed the importance of examining the agreement between clinical measurements with protein measurements from high throughput platforms. This thesis provides tools for and perspectives of the current status and future directions of OMICs research in epidemiology.

2 MEASUREMENT METHODOLOGY AND HIGH DIMENSIONALITY

Platforms such as SOMAscan and Metabolon provide quantification of more than a 1000 proteins and metabolites, respectively. However, these platforms come with their own specific shortcomings. First, the specificity and sensitivity of the metabolite or protein measurement differs depending on the chemical properties and nature of the biomolecule. For example, binding affinity of SOMAscan's aptamers are reportedly poor with proteins with a neutral charge or those with large sizes. Second, comparing and validating measurements with "golden" standard methods have been, at least partially, neglected as is evident from **chapter 2**. The appealing prospect of measuring a large number of metabolites and proteins should not detract from validating those measurements. Third, data should be double checked for post-processing errors during the annotation of the metabolites or proteins. The possibility of human errors and software errors occurring during this process is frequently ignored. Internal validations and simulations should be an essential element of the data integration and analysis process. In addition, improvement in measurement technology and methodology are needed to enable consistent measurements of complex metabolites or proteins. Furthermore, the number of detectable metabolites and proteins is projected to increase substantially. Therefore, in addition to validating the data using golden standard methods, researchers should validate findings use their own knowledge of the chemical properties and biochemical pathways of the metabolites and proteins related to their research question. Moreover, they must keep in mind the characteristics of the population of interest used in the study, since these may confound or affect the OMICs measurements. In conclusion, researchers must be aware of the strengths and weaknesses of the selected OMICs platforms used to produce their data and consider these when interpreting the results.

3 STATISTICAL CHALLENGES AND SOLUTIONS

3.1 The N<P problem

The merit of large-scale data from high throughput OMICs acts as a double-edged sword during the statistical analysis due to the multiple testing problem. For example, if 1000 comparisons are made between metabolites and a trait, the number of false positive results given a P-value threshold of 0.05 is 50 associations. This issue is commonly addressed by reducing the P-value threshold. For example, the Bonferroni method divides the nominal P-value by the number of independent measurements. Thus, studies must have an appropriately large sample size to assess associations between traits and a large number of measurements. Ideally the number of individuals in a study should be larger than the number biochemical variables. Otherwise, the analysis would suffer from N<P problem, wherein the number of individuals (N) in the study is less than the number of predictors (P). N<P leads to bias in the analysis results as well as reduced reliability and reproducibility of the results. In the case of epidemiological studies, it is essential to have a sufficient number of individuals per outcome event. Otherwise, etiological results may lack the necessary power to confidently report any findings and prediction models may be poor, overfitted, and not generalizable (1). Therefore, sample size considerations are crucial when conducting epidemiological studies using large OMICs data. The number of cohorts with OMICs measurements across the globe has increased greatly in recent years. To name a few notable examples, Nightingale measurements are available in the UK biobank cohort (n ~ 500,000) (2) and Metabolon is planned for the Million Veteran Program (n ~ 900,000) (3). Several consortia focusing on OMICs have also been established such as the biomolecular resources and research infrastructure (BBMRI) consortium (4, 5), BBMRI-ERIC (6), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (7, 8). These large studies and collaborative efforts provide several advantages to overcome the limitations of large OMICs data. They provide a solution for the N<P issues by combining their data and results to achieve larger power, similar to the work in **chapter 6**. Collaborative studies provide an additional benefit by enabling the reproduction and validation of findings in different populations. Meta-analyses of genome-wide association studies at the beginning of this century were in that sense pioneering, as they performed rigorous replication and validation studies. Lack of reproducibility in scientific research remains a persistent issue and OMICs research is no different. Thus, validation of prediction models and performing meta-analyses, increases the confidence in the research findings far more than single studies (9) as demonstrated in **chapter 7**. In addition, even significant findings in a specific population or ethnicity may not be generalizable to different populations and ethnicities. Hence, OMICs research should aim for the identification, comprehension, and dissolution of these disparities. To summarize, sample size considerations are critical for OMICs studies and collaborative research is crucial to not only overcome the N>P limitations but also to improve the quality and reproducibility of the findings.

3.2 Imputation of Missing Data

As we have shown in **chapter 3**, missing values are common and remain an issue in OMICs data. Complete case analysis, in which analysis is conducted on only individuals without any missing values, can be an appropriate method but also has its disadvantages (10). As discussed earlier, maximizing the sample size for analysis in OMICs is crucial. Thus, imputing missing values is an important method to reach full sample size. However, handling missing values in metabolomic studies requires careful consideration (11) as popular methods tend to be inadequate as they may lead to biased results (12, 13). This is the case for a common method for dealing with missing values in which metabolites with missingness above a certain percentage are excluded.

Alternatively, metabolites with missing values are imputed with a single numerical value, such as the mean or half the minimum value for each respective metabolite. However, this removes valuable data, and does not consider the various reasons for missingness in metabolomic data or the different characteristics of the measured metabolite groups (12, 13). For example, we recommend in **chapter 3** that xenobiotics should be imputed with zero as they are likely truly missing. For other metabolite groups, multiple imputation and k-nearest neighbor imputation demonstrated better performance with reduced bias in the analysis (12, 13). These methods are not perfect, and their performance dwindles if the sample sizes are small with high level of missingness. Overall, all these factors must be considered for metabolomics studies.

Another issue with handling missing values in metabolomic studies is that metabolomics research lacks a uniform guideline for imputation methodologies. Thus, selecting and applying appropriate imputation methods can be challenging. Multiple imputation may be most appropriate because of reduced potential bias but are thought to be computationally and statistically challenging (13). However, recent developments have overcome many of the issues associated with multiple imputation methods. Programing packages and plugins have become readily available with full tutorials on their usage, making it easier for researchers to apply. These include resources tailored for imputing missing values in metabolomic data (for example the script presented in **chapter 3**). These software tools coupled with the rapid evolution of computer processors and hardware have facilitated the application of imputation methods on large data. Further recommendations to reduce the complexity and computational time of multiple imputation methods is selecting a reasonable number of imputations for the study at hand. For example, 5 to 10 imputed sets can be sufficient in the presence of a moderate degree of missingness (14). Another recommendation is the selection of a small number of biologically relevant "auxiliary" variables (i.e., age, sex, body mass index, etc.) to impute the missing values for the metabolites instead of using the full dataset (13).

Moving forward, guidelines—such as the Metabolomics Standards Initiative (15)—should be expanded upon to provide a uniform primer for imputing missing data using optimized methodologies for metabolomics. Achieving this would require the collaboration and agreement of the metabolomics research community and is definitely something worth pursuing.

4 REPRESENTING AND INTERPRETING METABOLOMICS AND PROTEOMICS RESULTS

The amount of data measured by OMICs platforms raises another question: how do we represent and interpret the vast amount of data from the analysis? Typical results of epidemiological studies are provided in tables and static figures within the main body of the manuscript. However, when reporting large numbers of tests for every measured metabolite, protein, or genetic variation, it becomes challenging to include all the results in a single table or figure. Moreover, scientific journals typically limit the number of tables and figures allowed in the main text. Providing these results as supplementary materials creates dozens of large, sometimes overwhelming tables and figures, which are often overlooked. This leads to selecting and reporting a handful of metabolites, proteins, or single nucleotide polymorphisms (SNPs) in the main body of scientific papers. Moreover, OMICs measurements, especially metabolites and proteins, are highly correlated and interconnected by common chemical pathways. Therefore, reporting on a select few compounds fails to capture the true depth of their biological implications, which subsequently can lead to bias in reporting.

Reporting of OMICs data can be improved in a number of ways. First, the full analysis results can be published on an online website, such as an accessible online database or interactive webpage viewed directly from an internet browser. This enables easy viewing and interaction with large tables and complex figures. In addition, a website dedicated to the results enables the inclusion of all the statistical estimates from the analysis. Examples of these sites already exist, such as the Metabolomics GWAS Server (16, 17), the Metabolome website for metabolites correlated with age (18, 19), and our website for the results regarding metabolites associated with hepatic triglyceride content, described in **chapter 5**. Other alternatives are central web databases and online analysis suites for uploading OMICs data and performing additional analyses. Examples of these resources include MetaboLights (20) and OpenGWAS (21). Second, analytical methods should be available that capture the extent of correlations and association within OMICs data. As demonstrated in **chapter 5**, methods such as gaussian graphical modeling (GGM) and genome-scale metabolic models (GSMM) provide unique insight into pathways and reactions connecting the OMICs data and the analysis results (22). Third, dynamic web figures and plots can be generated to be interactive instead of static images. This facilitates the navigation and interaction with complex results while providing the ability for both wide and pinpoint visualization.

With software and programming advancements, the tools to enable the production of such interactive figures and websites have become available for use by researchers. In addition, some tools have been designed specifically for the online representation of OMICs data. One such tool is PheWeb for the visualization and exploration of genomic study results (23). Moreover, interactive GGM and GSMM networks can be produced by common and free software. Thus, it is now feasible for researchers to create interactive figures and networks that allow full and thorough exploration and investigation of OMICs studies.

5 XENOBIOTICS, A KEY FOR EXPOSOME RESEARCH?

Exposome, the study of the effects of external environmental and lifestyle exposures on individuals' health, has gained steady momentum in recent years. Indeed, large consortia have been founded dedicated to exposome research, such as the exposome-NL consortium (24) in the Netherlands. However, one persistent limitation of these studies is the poor availability of real-life exposome measurements for the general population. Metabolomics can be a solution for this issue. Indeed, some interesting exposome variables can be detected in the xenobiotic metabolite measurements found in metabolomic platforms such as Metabolon. For example, some environmental contaminants such as pollutants in the air, soil, or water may be measured in body fluids, such as the forever chemicals (PFAS) exposure levels as demonstrated in **chapter 7**. In addition, xenobiotic metabolites such as cotinine metabolites reflect smoke exposure and other lifestyle habits. Besides environmental exposures, xenobiotic metabolites may reflect components from diet, cosmetics, and medications. Diet related metabolites can be used in unique ways in epidemiological studies. For example, some metabolites have been linked with dietary patterns such as the intake of fish and bread. Several studies have explored such biomarkers as quantitative information for dietary intake and the validation of food frequency questionnaires (FFQs) (25, 26). These FFQs can be affected by participants misremembering what they ate over short or long periods of time. This issue is called differential measurement error bias and is commonly referred to as "recall bias" (27). Therefore, quantitative diet information via metabolite measurements can address recall bias and complement the FFQs. However, achieving this goal remains a challenge that requires the identification and validation of strong and robust biomarkers for a wide number diets and food sources (26). The same principle of validating patient questionnaires using metabolites can potentially

be applied to verify other lifestyle factors such as the aforementioned tobacco smoking via cotinine metabolites (28).

Xenobiotic metabolites from medications can also provide unique epidemiological insights. Common ways to obtain medication data is from prescription and dispensing data from hospitals or pharmacies, or from the beforementioned self-reported patient questionnaires. However, this data could be scarce for the general population or not accessible for research purposes. In addition, some drugs can be obtained over the counter without a prescription and do not leave patient specific data traces. In these cases, metabolomic measurements can be useful if the medication related xenobiotic metabolites can be quantified. Nonsteroidal anti-inflammatory drugs (NSAIDs) are an example of an over-the-counter drug class that does not require prescription and are difficult to trace. This is an issue when attempting to study the negative health impact of NSAIDs overuse. Indeed, NSAIDs have been found to be associated with increased risk of heart failure (29)—particularly in hypertensive patients (30)—and increased risk of gastrointestinal bleeding (31, 32). The same principle can also be coupled with randomized controlled trials (RCTs) to collect metabolomic quantitative data regarding the patient's medication compliance. Therefore, using xenobiotic metabolites for NSAIDs and other medications can provide quantitative data of their usage in the general population. In turn, these xenobiotic measurements can be used to address various epidemiological questions such as improving the estimation of the population use of NSAIDs—or other medications—and their association with negative health outcomes or mortality.

One limitation that should be noted for the use of metabolomics to trace and quantify medication related metabolites is the half-life of metabolites in the human samples, i.e., the period of time it takes a substance (metabolite) level to decrease to half its initial concentration (33). A metabolite with a short half-life is eliminated from the body too quickly and can be difficult to measure. One possible work around for this issue is to obtain and measure multiple samples to capture metabolite levels at different time points.

As demonstrated by these examples, the quantification of xenobiotic metabolites related to environmental contamination, diet, or medication could provide tools to answer exposome related epidemiological questions. Hence, metabolomics may be a key for the expansion of exposome research.

6 THE VALUE OF CROSSING AND INTEGRATING MULTIPLE OMICS

The human biological system is incredibly complex and sophisticated. Thanks to technological advancements, OMICs have enabled the comprehensive study of different layers of this human system. OMICs studies usually focus on a single layer; however, these layers are interconnected and actively interacting. Although the genetic sequence is largely static and conserved after conception, massive diversity in gene expression occurs from epigenetic regulation. Moreover, gene expression differs in various body tissues and the degree of expression differs over time. As a consequence, the diversity and levels of proteins and metabolites in different tissues also differ over time.

This level complexity is disregarded when focusing on a single OMICs layer. Indeed, despite the important findings from genome wide association studies, examining DNA sequence data and SNPs alone is currently incapable of explain the full heritability of diseases, referred to as the missing heritability problem (34). Likewise, focusing on OMICs data such as metabolomics alone ignores genetic factors that may affect metabolism and metabolite levels.

The integration of two or more OMICs, i.e., multi-OMICs, can be a powerful approach that combines and reinforces the unique features of separate OMICs data. Indeed, multi-OMICs can expand on etiological findings by providing better understanding of disease pathophysiology. For example, the integration of genomic and metabolomic data as presented in **chapter 6,** has also been used for NAFLD research (35), and to create an atlas of genetic influences on blood metabolites (16). Another valuable addition to a multi-OMICs study is the use of transcriptomics data. Transcriptomics can be further applied to assess the dynamic expression of the genome in different cell lines in the body. Crossing over the results from genomics and transcriptomics with metabolomics or proteomics can form a network pinpointing the location and time points for gene expression and link it to the levels of proteins and metabolites. Beyond these endogenous layers, the beforementioned external xenobiotics and exposome can also be added to this network. Thus, the possible associations of environmental and lifestyle exposures with the endogenous factors can be considered. In this sense, crossing OMICs not only combines their unique features, but also compensates their individual downsides and reinforces their strengths. Therefore, integrating multi-OMICs data is an important up and coming field for understanding complex pathophysiology of human phenotypes and diseases.

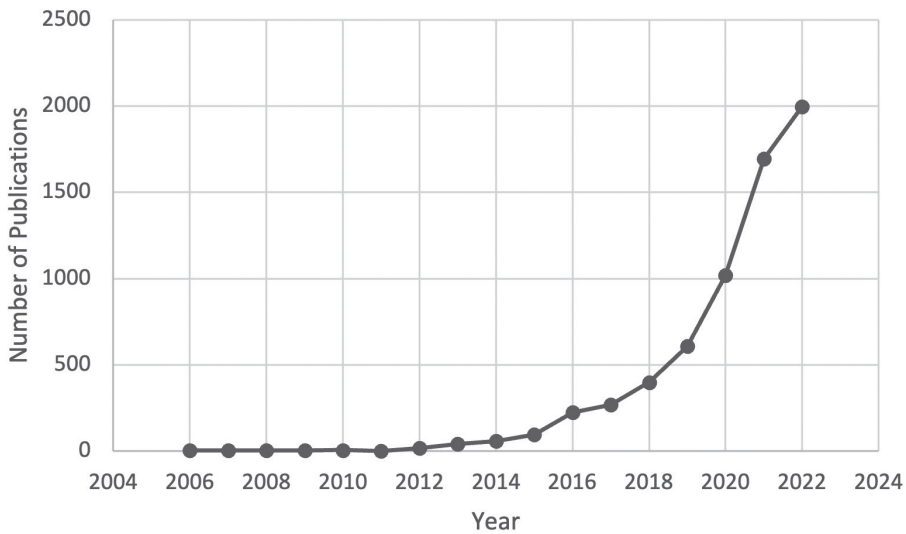
Crossing over multi-OMICs can also provide insights into causal associations. One way to achieve this is by using genomics in combination with one or more OMICs data. Since genetic variations are conserved since conception and are not affected by confounders, it is possible to use genomics to infer causality, using statistical methodologies such as Mendelian Randomization (MR) (36). On the other hand, metabolomics and proteomics are not readily applicable to assess causality in associations with diseases. However, genomics combined with other single OMICs can possibly enable inferring causality between the different single OMICs and outcomes of interest. In an MR analysis SNPs can be selected that are associated with a specific outcome and relevant metabolites or proteins biomarkers. These can then be used to estimate the causal effect of the biomarkers on the outcome. Thus, providing a method to avoid confounding if the assumptions required for causal inference in MR are maintained. A hypothetical example would be for venous thrombosis research. The search for biomarkers which are associated or predictive for VTE remains challenging (37, 38). Multi-OMICs can be utilized by using known and unknown genomic associations with VTE and its associated metabolites and proteins. Analysis of these OMICs could potentially assist in identifying novel biomarkers or causal associations related to VTE (by way of MR or other causal inference methodologies). To date, several multi-OMICs MR studies have been used to identify potential drug targets (39), inferred causality between the metabolome and obesity (40), and identified causal links between carnitine and systolic blood pressure (41).

7 MULTI-OMICS AND THE (NOT SO) LONG AND WINDING ROAD

We have described the merits of multi-OMICs and provided examples which demonstrate their potential, such as creating complex OMICs atlases for deeper etiological understanding of diseases (16, 42, 43) and in identifying causal associations (39-42). But what are the challenges facing multi-OMICs? What is needed for it to grow? How far are we in this upcoming field? For starters, the big challenge of multi-OMICs data is the sheer amount of data to combine and study. One way to address this challenge is to use the aforementioned analytical methods to construct a multi-OMICs network for the relevant biological pathways related to the outcome of interest. This selection can be done using statistical methods or prior findings from epidemiological and single OMICs studies.

Multi-OMICs studies are currently being conducted more frequently (Figure 1). However, for multi-OMICs to blossom, it would first require further advancements in the separate OMICs fields. These advancements must aim to improve the reliability and reproducibility of measurements whilst reducing costs. For genomics, we reiterate the importance of studying copy number variations, structural variations, and other variations in addition to SNPs. This may help to address the missing heritability problem and may shed light on new mechanisms underlying pathophysiology of diseases. In addition, new and efficient methods are needed to integrate and analyze the complex OMICs data. Open-source software and plugins for statistical software and programming languages, such as R and Python, can streamline these methodologies and simplify their use for multi-OMICs research. The subsequent results and findings would greatly benefit from dynamic and interactive representation to enable the full exploration by other researchers. Indeed, network analysis and interactive web tools can display multi-OMICs data and their associations efficiently. In addition, new analysis methods have been explored to process and analyze multi-OMICs data. One such method is machine learning (ML). ML techniques are increasingly applied in epidemiological and OMICs research. ML is a powerful technique for the analysis of large data and can be used for prediction modeling and even causal effect estimation (44). However, ML may suffer from several pitfalls. For one, the parameters and steps silently undertaken by the ML algorithm, to automatically learn from the data, may be nontransparent and complicated (45, 46). Moreover, ML inherently uses some level of randomness which may affect its performance and specific outcome. Changing the randomness parameter (also known as the random seed) may lead to inflated estimates of the model (45, 47) Another aspect is that ML methodologies can also be affected from the same issues as non-ML methods, such confounding, overfitting, and bias (45, 47). These aspects can be challenging when verifying and reproducing the results from ML models (45, 47). In addition, some studies reported that, for clinical prediction models, ML showed no performance benefits when compared to logistical regression (48). These problems have been addressed by combining epidemiological principles and statistical frameworks with ML and by selecting appropriate ML algorithms and properly applying them (44-46). Indeed, efforts are underway to provide guidelines for to address the aforementioned issues and help guide the development and reporting of ML prediction models (46). These aspects and solutions will be applicable for multi-OMICs as well and can enable the use of robust ML methods in OMICs and multi-OMICs research.

Figure 1: Search results of publications using the term "multi-OMICs" in PubMed from 2006 to November 2022 (49)



Ultimately, the potential of multi-OMICs research is promising and the number of studies incorporating multiple OMICs is rapidly increasing. Furthermore, statistical methodologies, such as ML, are improving as well, further enabling multi-OMICs analyses. Overall, the challenges facing multi-OMICs are being addressed and will likely be steadily resolved. The road to prevalent multi-OMICs epidemiological research appears shorter than ever before.

8 OMICS, CLINICAL CARE, AND BEYOND

OMICs and multi-OMICs are powerful tools for dissecting the pathophysiology and etiology of diseases. How can these findings be incorporated into clinical care? Physicians and clinicians rely on their medical knowledge and experience for prognosis, diagnosis, and treatment of patient's medical conditions. Based on this knowledge, they may request a screening for specific biomarkers with a strong indication for a specific outcome to verify their diagnosis. For example, c-reactive protein (CRP) is a strong biomarker for inflammation and used as a diagnostic marker of infection and inflammation (50). D-dimer, small protein fragments of fibrin, is a strong diagnostic biomarker for venous thrombosis events (51). D-dimer has a high sensitivity for the identification VTE patients but is noted to having low specificity to identify patients without VTE (52). Prohormone brain natriuretic peptide (pro-BNP) is an example of a peptide that has become regularly measured for the diagnosis, particularly in the emergency room, and prognosis of heart failure (53-55). Physicians also rely on the patient's personal and family history to order screening of specific disease biomarkers. For example, screening for the BRCA1 and BRCA2 genes has become common procedure to identify the risk of breast cancer and ovarian cancer (56). Another example is prenatal and newborn genetic screening. These tests have become routine clinical procedures to identify possible treatable but severe diseases that require swift and early action for new born babies (57). Newborn genetic screenings checks for diseases such as thyroid disorder, blood disorders, a range of metabolic disorders, and several other diseases (58). These examples show how some OMICs findings, especially genomics, have reached and enhanced some areas of routine clinical care (59, 60). In order to be implemented in clinical practice, identified biomarkers for diseases must show a robust and specific association with a disease outcome, show a strong utility for prognosis and diagnosis of an outcome, and should be readily available and easily measured. Subsequently, this may lead to the incorporation of the biomarker measurement to routine clinical application, as what was done for pro-BNP. However, challenges still remain, and progress is slower than expected for the integration of OMICs biomarkers in the clinic. As a consequence, the integration of these biomarkers into clinical use has been slow. For example despite strong evidence linking genetic variation with increased disease risk, it has taken a 25-30 year period between discovery and clinical implementation of genetic markers such BRCA1 (61) and nearly 10 years for incorporation and standardization of pro-BNP measurement in clinical care (62).

Genetic screening for patients remains expensive and time consuming. This adds another layer of complication due to limited financial coverage of these tests by health insurance companies. This financial burden discourages physicians and patients to request such tests. In addition, results for the analysis are often not easily readable by the clinician or the risk probability from the tests are not sufficient to determine a treatment course. "Why request an expensive and time-consuming test that does not add value to the standard treatment plan? Is OMICs research truly useful for clinical care?" These concerns and issues must be addressed when thinking of the future value of OMICs to clinical setting. Metabolomics and proteomics research must learn from the lesson of genomics and other successful biomarkers (such as pro-BNP) to prove their value and to expedite their integration in the clinical setting. We will discuss three potential points that can aid OMICs to reach clinical care.

First, the measurements from OMICs platform can provide better clinical relevance if they report the absolute concentration of the biomarkers instead of relative values. One of the tradeoffs of most untargeted methodologies is between the ability to measure large amounts of biomarkers and obtaining absolute quantification. One of the reasons for this is the nature of the methods and platforms used to a attain those measurements (i.e., mass spectrometry and aptamer-based methods). The large number of measured biomarkers and relative concentrations from these platforms are indeed useful and beneficial to address research questions. However, for clinical use these relative measurements can be difficult to interpret when deciding a treatment plan. This is particularly true when compared to standardized and routine clinical biomarker measurements. Of course, attaining absolute quantification is possible from some OMICs instruments and platforms, such as the Nightingale platform. However, moving forward, OMICs should aim to provide absolute concentrations on large scales. For existing absolute quantification platforms this would require expanding the range of measurable biomarkers. For relative quantification platforms this requires finding solutions such as converting relative values to absolute concentrations by including reference samples or implementing other methodologies. Alternatively, future technologies can lead the way for novel large scale untargeted absolute quantification platforms. Achieving these goals will ultimately aid in translating the findings from OMICs research into clinical relevancy.

Second, as discussed earlier, identification of causal associations from single or multi-OMICs research could be the key to finding novel and important disease biomarkers. This can be further expanded to identify potential protein and metabolite drug targets. Previous studies have successfully used MR to identify ACE2 and IFNAR2 proteins as potential drug targets for severe COVID-19 patients (63). This was in line with findings from a large randomized controlled trial (RCTs) (64). This study also showcases other benefits of MR and OMICs studies. Compared with an RCT, an MR study is much cheaper, less time consuming, and faces fewer ethical implications. In addition, many RCTs fail at the crucial phase 3 stage (65), thus wasting time, money, and potentially negatively affecting the wellbeing of patients included in the RCT. One of the reasons for this high rate of failure is designing drugs for a target without sufficient causal evidence (65, 66). An example of this is an MR study that concluded, despite previous expectations, that CETP and CETP inhibition showed no causal association with reducing cardiovascular disease risk and therefore was not a suitable drug target for prevention of CVD (67). This illustrates how failed RCTs, like the ones who targeted CETP, could have been prevented if genetic and OMICs findings were initially applied. Indeed, OMICs and multi-OMICs evidence could complement epidemiological studies and provide insight for the design of RCTs (68-70). Thus, researching this combination would be beneficial as it can potentially increase the probability of success of RCTs, while potentially reducing financial cost and reducing patient burden.

Third, biomarkers are rare, which in isolation provide strong evidence for the prognosis and diagnosis of a disease or health outcome. An alternative method may be to combine several biomarkers to generate a single score to diagnose or predict a disease. In genomics, this already has become commonplace by means of polygenic risk scores (71). A similar method was used in **chapter 4**, in which hundreds of metabolites were combined to provide a “metabolomic age”. Similarly, a combination of metabolites, proteins, and genetic variations may be used to identify disease specific profiles (72). For example, these profiles could aid in the identification of patients with a high risk of cardiometabolic disease despite exhibiting a normal weight—which is typically associated with a low risk of cardiometabolic disease. Conversely, it can also identify overweight individuals who are biologically at low risk of cardiometabolic disease. These two groups are sometimes referred to as exhibiting unfavorable and favorable adiposity respectively, and have been reported to be linked with specific SNPs (73, 74) and specific SNP-metabolomic profiles

(75). Identifying these individuals without OMICs data may be challenging for physicians. In an ideal scenario, a panel of a different OMICs results associated with favorable and unfavorable adiposity is developed. The measurements from these panels can be used for the diagnosis or risk assessment of cardiometabolic disease on an OMICs level. These types of panels can also help in reducing cost and time to produce results for clinical use.

In conclusion, OMICs research is important for understanding and disentangling disease pathophysiology, discovering novel associations, revealing effects of exposures, identifying causal pathways, bridging the gap between the roles of nature and nurture, and enhance public health and clinical care. With the continuous expansion of single and multi-OMICs studies and rapid technological advancements, it is not farfetched to expect more impactful findings in the near future. Ideally, the time between discovery to clinical application will be shortened as well.

9 REFERENCES

1. Steyerberg EW. Clinical Prediction Models. 2nd ed. Cham, Switzerland: Springer International Publishing; 2019 2019.
2. UK Biobank adds the first tranche of data from a study into circulating metabolomic biomarkers to its biomedical database 2022 [updated 2022/08/26/. Available from: https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/uk-biobank-adds-the-first-tranche-of-data-from-a-study-into-circulating-metabolomic-biomarkers-to-its-biomedical-database.
3. Metabolon. Metabolon to Provide Metabolomic Profiling for the US Veterans Administration Million Veteran Program 2022 [updated 2022/08/09/. Available from: https://www.metabolon.com/news/metabolon-to-provide-metabolomic-profiling-for-the-us-veterans-administration-million-veteran-program.
4. BBMRI | BBMRI 2022 [updated 2022/11/01/. Available from: https://www.bbmri.nl.
5. Onderwater GLJ, Ligthart L, Bot M, Demirkan A, Fu J, van der Kallen CJH, et al. Large-scale plasma metabolome analysis reveals alterations in HDL metabolism in migraine. 2019;92(16):e1899-e911.
6. Litton J-E. Launch of an Infrastructure for Health Research: BBMRI-ERIC. 2018;16(3):233-41.
7. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet. 2009;2(1):73-80.
8. CHARGE Consortium Wiki 2022 [updated 2022/10/17/. Available from: http://depts.washington.edu/chargeco/wiki/Main_Page.
9. Arya S, Kaji AH, Boermeester MA. PRISMA Reporting Guidelines for Meta-analyses and Systematic Reviews. JAMA Surgery. 2021;156(8):789-90.
10. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. International journal of epidemiology. 2019;48(4):1294-304.
11. Hrydziuszko O, Viant MR. Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. Metabolomics. 2012;8(1):161-74.
12. Faquih T, van Smeden M, Luo J, le Cessie S, Kastenmüller G, Krumsiek J, et al. A Workflow for Missing Values Imputation of Untargeted Metabolomics Data. Metabolites. 2020;10(12).
13. Do KT, Wahl S, Raffler J, Molnos S, Laimighofer M, Adamski J, et al. Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. Metabolomics. 2018;14(10):128.
14. van Buuren S. Flexible Imputation of Missing Data. Second Edition. Boca Raton, FL.: CRC Press; 2018.
15. Fiehn O, Robertson D, Griffin J, van der Werf M, Nikolau B, Morrison N, et al. The metabolomics standards initiative (MSI). Metabolomics. 2007;3(3):175-8.
16. Shin S-Y, Fauman EB, Petersen A-K, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nature Genetics. 2014;46(6):543-50.
17. Metabolomics GWAS Server. 2022.
18. Kong SW. Metabolites correlated with age 2021 [updated 2021/05/03/. Available from: https://tom.tch.harvard.edu/supples/metabolome/age-correlation.htm.
19. Kong SW, Hernandez-Ferrer C. Assessment of coverage for endogenous metabolites and exogenous chemical compounds using an untargeted metabolomics platform. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing. 2020;25:587-98.
20. The MetaboLights Team MetaboLights 2022 [updated 2022/08/26/. Available from: https://www.ebi.ac.uk/metabolights/studies.
21. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et al. The MRC IEU OpenGWAS data infrastructure. 2020:2020.08.10.244293.
22. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. BMC Systems Biology. 2011;5(1):21.

23. Gagliano Taliun SA, VandeHaar P, Boughton AP, Welch RP, Taliun D, Schmidt EM, et al. Exploring and visualizing large-scale genetic associations by using PheWeb. Nat Genet. 2020;52(6):550-2.
24. Exposome-NL. Exposome-NL. 2022.
25. Arab L, Tseng CH, Ang A, Jardack P. Validity of a multipass, web-based, 24-hour self-administered recall for assessment of total energy intake in blacks and whites. American journal of epidemiology. 2011;174(11):1256-65.
26. Gibbons H, Brennan L. Metabolomics as a tool in the identification of dietary biomarkers. Proceedings of the Nutrition Society. 2017;76(1):42-53.
27. Sheikh MA, Abelsen B, Olsen JA. Differential Recall Bias, Intermediate Confounding, and Mediation Analysis in Life Course Epidemiology: An Analytic Framework with Empirical Example. Frontiers in psychology. 2016;7:1828.
28. Folpmers S, Mook-Kanamori DO, de Mutsert R, Rosendaal FR, Willems van Dijk K, van Heemst D, et al. Agreement between nicotine metabolites in blood and self-reported smoking status: The Netherlands Epidemiology of Obesity study. Addictive behaviors reports. 2022;16:100457.
29. Schjerning AM, McGettigan P, Gislason G. Cardiovascular effects and safety of (non-aspirin) NSAIDs. Nature reviews Cardiology. 2020;17(9):574-84.
30. Bavry AA, Khaliq A, Gong Y, Handberg EM, Cooper-DeHoff RM, Pepine CJ. Harmful Effects of NSAIDs among Patients with Hypertension and Coronary Artery Disease. The American Journal of Medicine. 2011;124(7):614-20.
31. Straube S, Tramèr MR, Moore RA, Derry S, McQuay HJ. Mortality with upper gastrointestinal bleeding and perforation: effects of time and NSAID use. BMC Gastroenterology. 2009;9(1):41.
32. Bedene A, van Dorp ELA, Rosendaal FR, Dahan A, Lijfering WM. Risk of drug-related upper gastrointestinal bleeding in the total population of the Netherlands: a time-trend analysis. 2022;9(1):e000733.
33. Hallare J, Gerriets V. Half Life. 2022 2022 Jun 23. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing. Available from: https://www.ncbi.nlm.nih.gov/books/NBK554498/.
34. Génin E. Missing heritability of complex diseases: case solved? Human Genetics. 2020;139(1):103-13.
35. Sliz E, Sebert S, Würtz P, Kangas AJ, Soininen P, Lehtimäki T, et al. NAFLD risk alleles in PNPLA3, TM6SF2, GCKR and LYPLAL1 show divergent metabolic effects. Human molecular genetics. 2018;27(12):2214-23.
36. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
37. Goldenberg NA, Everett AD, Graham D, Bernard TJ, Nowak-Göttl U. Proteomic and other mass spectrometry based "omics" biomarker discovery and validation in pediatric venous thromboembolism and arterial ischemic stroke: Current state, unmet needs, and future directions. PROTEOMICS – Clinical Applications. 2014;8(11-12):828-36.
38. Cushman M, Barnes GD, Creager MA, Diaz JA, Henke PK, Machlus KR, et al. Venous thromboembolism research priorities: A scientific statement from the American Heart Association and the International Society on Thrombosis and Haemostasis. Research and practice in thrombosis and haemostasis. 2020;4(5):714-21.
39. Russell C, Rahman A, Mohammed AR. Application of genomics, proteomics and metabolomics in drug discovery, development and clinic. Therapeutic delivery. 2013;4(3):395-413.
40. Hsu YH, Astley CM, Cole JB, Vedantam S, Mercader JM, Metspalu A, et al. Integrating untargeted metabolomics, genetically informed causal inference, and pathway enrichment to define the obesity metabolome. International journal of obesity (2005). 2020;44(7):1596-606.
41. Richard MA, Lupo PJ, Zachariah JP. Causal Inference of Carnitine on Blood Pressure and potential mediation by uric acid: A mendelian randomization analysis. Int J Cardiol Cardiovasc Risk Prev. 2021;11:200120.
42. Surendran P, Stewart ID, Au Yeung VPW, Pietzner M, Raffler J, Wörheide MA, et al. Rare and common genetic determinants of metabolic individuality and their effects on human health. Nature Medicine. 2022;28(11):2321-32.

43. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73-9.
44. Balzer LB, Petersen ML. Invited Commentary: Machine Learning in Causal Inference-How Do I Love Thee? Let Me Count the Ways. American journal of epidemiology. 2021;190(8):1483-7.
45. Beam AL, Manrai AK, Ghassemi M. Challenges to the Reproducibility of Machine Learning Models in Health Care. Jama. 2020;323(4):305-6.
46. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021;11(7):e048008.
47. McDermott MBA, Wang S, Marinsek N, Ranganath R, Foschini L, Ghassemi M. Reproducibility in machine learning for health research: Still a ways to go. Science Translational Medicine. 2021;13(586):eabb1655.
48. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of Clinical Epidemiology. 2019;110:12-22.
49. PubMed. multi-omics - Search Results - PubMed 2022 [updated 2022/11/14/. Available from: https://pubmed.ncbi.nlm.nih.gov/?term=multi-omics&filter=years.2006-2022&show_snippets=off.
50. Sproston NR, Ashworth JJ. Role of C-Reactive Protein at Sites of Inflammation and Infection. Frontiers in immunology. 2018;9:754.
51. Kearon C, de Wit K, Parpia S, Schulman S, Spencer FA, Sharma S, et al. Diagnosis of deep vein thrombosis with D-dimer adjusted to clinical probability: prospective diagnostic management study. 2022;376:e067378.
52. Pulivarthi S, Gurram MK. Effectiveness of d-dimer as a screening test for venous thromboembolism: an update. North American journal of medical sciences. 2014;6(10):491-9.
53. Maisel AS, Krishnaswamy P, Nowak RM, McCord J, Hollander JE, Duc P, et al. Rapid Measurement of B-Type Natriuretic Peptide in the Emergency Diagnosis of Heart Failure. 2002;347(3):161-7.
54. Nakagawa O, Ogawa Y, Itoh H, Suga S, Komatsu Y, Kishimoto I, et al. Rapid transcriptional activation and early mRNA turnover of brain natriuretic peptide in cardiocyte hypertrophy. Evidence for brain natriuretic peptide as an "emergency" cardiac hormone against ventricular overload. The Journal of Clinical Investigation. 1995;96(3):1280-7.
55. Sudoh T, Kangawa K, Minamino N, Matsuo HJN. A new natriuretic peptide in porcine brain. 1988;332(6159):78-81.
56. Toland AE, Forman A, Couch FJ, Culver JO, Eccles DM, Foulkes WD, et al. Clinical testing of BRCA1 and BRCA2: a worldwide snapshot of technological practices. npj Genomic Medicine. 2018;3(1):7.
57. Krstić N, Običan SG. Current landscape of prenatal genetic screening and testing. 2020;112(4):321-31.
58. Rijksinstituut voor Volksgezondheid en Milieu. Heel prick screening test | Prenatale en neonatale screeningen 2022 [updated 2022/09/20/. Available from: https://www.pns.nl/prenatal-and-newborn-screening/heel-prick.
59. Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Archives of Toxicology. 2011;85(1):5-17.
60. McCarthy JJ, McLeod HL, Ginsburg GS. Genomic Medicine: A Decade of Successes, Challenges, and Opportunities. Science Translational Medicine. 2013;5(189):189sr4-sr4.
61. Gibbs RA. The Human Genome Project changed everything. Nature Reviews Genetics. 2020;21(10):575-6.
62. Semenov AG, Feygina EE. Chapter One - Standardization of BNP and NT-proBNP Immunoassays in Light of the Diverse and Complex Nature of Circulating BNP-Related Peptides. In: Makowski GS, editor. Advances in Clinical Chemistry. 85: Elsevier; 2018. p. 1-30.
63. Gaziano L, Giambartolomei C, Pereira AC, Gaulton A, Posner DC, Swanson SA, et al. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. Nature Medicine. 2021;27(4):668-76.

64. Lopes RD, Macedo AVS, de Barros ESPGM, Moll-Bernardes RJ, Feldman A, D'Andréa Saba Arruda G, et al. Continuing versus suspending angiotensin-converting enzyme inhibitors and angiotensin receptor blockers: Impact on adverse outcomes in hospitalized patients with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)--The BRACE CORONA Trial. American heart journal. 2020;226:49-59.
65. Seruga B, Ocana A, Amir E, Tannock IF. Failures in Phase III: Causes and Consequences. Clinical cancer research : an official journal of the American Association for Cancer Research. 2015;21(20):4552-60.
66. Stewart DJ, Kurzrock R. Fool's gold, lost treasures, and the randomized clinical trial. BMC cancer. 2013;13:193.
67. Blauw LL, Li-Gao R, Noordam R, de Mutsert R, Trompet S, Berbée JFP, et al. CETP (Cholesteryl Ester Transfer Protein) Concentration: A Genome-Wide Association Study Followed by Mendelian Randomization on Coronary Artery Disease. Circulation Genomic and precision medicine. 2018;11(5):e002034.
68. Ference BA, Holmes MV, Smith GD. Using Mendelian Randomization to Improve the Design of Randomized Trials. Cold Spring Harbor perspectives in medicine. 2021;11(7).
69. Henry A, Gordillo-Marañón M, Finan C, Schmidt AF, Ferreira JP, Karra R, et al. Therapeutic Targets for Heart Failure Identified Using Proteomics and Mendelian Randomization. Circulation. 2022;145(16):1205-17.
70. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human molecular genetics. 2014;23(R1):R89-98.
71. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Medicine. 2020;12(1):44.
72. Suhre K, Zaghlool S. Connecting the epigenome, metabolome and proteome for a deeper understanding of disease. Journal of internal medicine. 2021;290(3):527-48.
73. Ji Y, Yiorkas AM, Frau F, Mook-Kanamori D, Staiger H, Thomas EL, et al. Genome-Wide and Abdominal MRI Data Provide Evidence That a Genetically Determined Favorable Adiposity Phenotype Is Characterized by Lower Ectopic Liver Fat and Lower Risk of Type 2 Diabetes, Heart Disease, and Hypertension. Diabetes. 2018;68(1):207-19.
74. Yaghootkar H, Lotta LA, Tyrrell J, Smit RAJ, Jones SE, Donnelly L, et al. Genetic Evidence for a Link Between Favorable Adiposity and Lower Risk of Type 2 Diabetes, Hypertension, and Heart Disease. Diabetes. 2016;65(8):2448-60.
75. Cirulli ET, Guo L, Leon Swisher C, Shah N, Huang L, Napier LA, et al. Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk. Cell Metabolism. 2019;29(2):488-500.e2.