Your blood may already know what illness comes next—long before symptoms appear, study finds

May 14 2026, by Pranjal Malewar

Credit: Image generated by the editorial team using AI for illustrative purposes.

Predicting who will develop common diseases is key to prevention, detection, and early treatment. Traditionally, clinicians have estimated

risk based on age, sex, laboratory results, and lifestyle factors. Although these classical indicators provide important information, they do not necessarily reflect the multifaceted biological mechanisms underlying disease development.

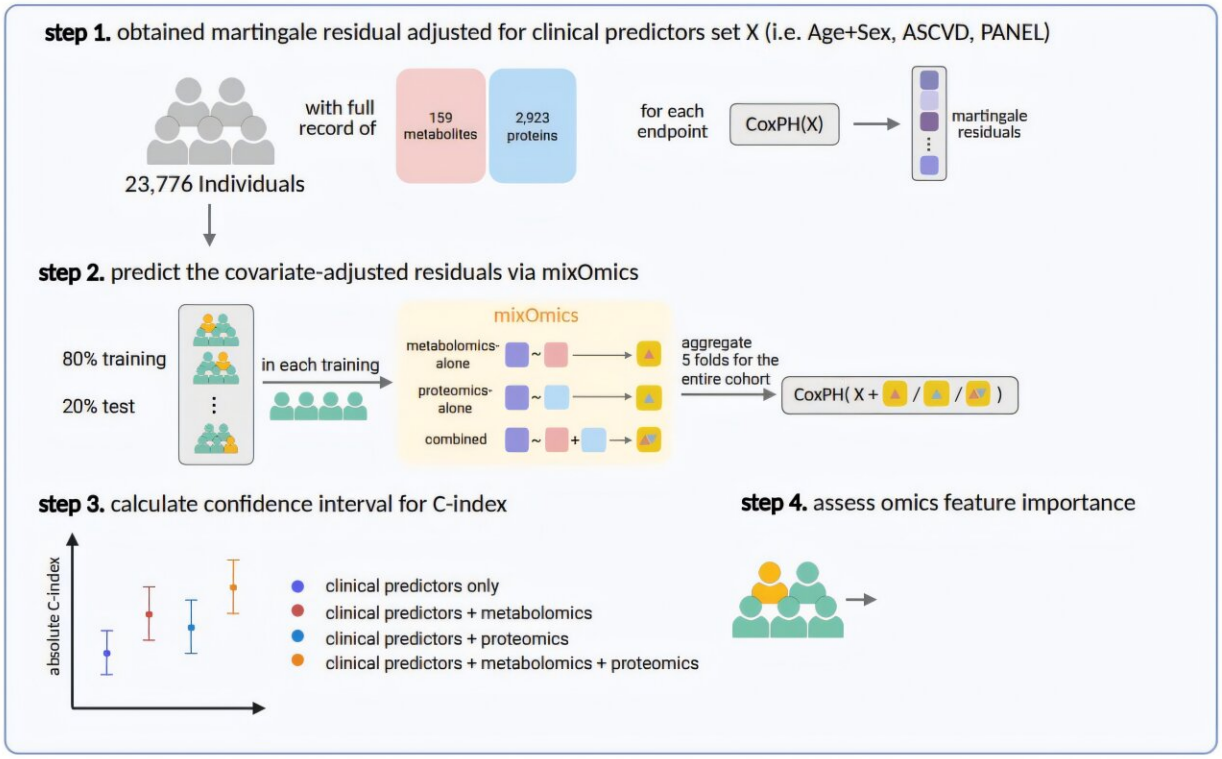
Researchers are now peering deeper into the hidden molecular world inside us. Multi-omics, an advanced biological analysis approach that integrates data from multiple "omes," such as genomics, transcriptomics, proteomics, and metabolomics, provides a comprehensive view of biological systems. It is like opening several windows at once. Each layer adds a clue, each clue adds a story.

Of the different types of omics technologies, proteomics (the study of proteins) and metabolomics (the study of metabolites) have been especially promising. Nevertheless, despite the advances achieved through large-scale approaches in these areas of research, studies integrating both remain scarce, and each new study is another step forward toward personalized medicine.

Publishing in *Nature Communications*, the researchers studied almost 24,000 people from the UK Biobank to see if adding detailed molecular information, like proteins (proteomics) and small molecules (metabolomics), could help predict disease risk.

They found that using this additional data improved prediction accuracy across all 17 diseases they studied, including cancers, heart problems, diabetes, brain disorders, and lung diseases. This worked better than relying only on standard clinical measures.

The authors noted, "To the best of our knowledge, our study is currently the largest to systematically evaluate contributions of both metabolomics and proteomics profiles in incident disease prediction."



Infographic summarizing the study design and workflow. Credit: *Nature Communications* (2026). DOI: 10.1038/s41467-026-73017-z

Metabolites show the bigger picture

Metabolomics provides inexpensive measures of circulating metabolites and can outperform traditional risk factors in predicting health. Proteomics is the quantification of proteins in blood and tissues to characterize disease biology that enhances risk prediction. Large-scale studies that combine both are rare.

Combined utilization of these complementary strategies may lead to improved disease prediction, allowing early diagnosis, precise prognosis,

and ultimately personalized treatment approaches.

Given this gap, researchers performed a large-scale analysis of 159 metabolites and 2,923 proteins from nearly 24,000 individuals in the UK Biobank (a key resource for grand-scale studies bridging molecular data with disease). To determine the added predictive value of metabolomics and proteomics data (versus conventional clinical variables), researchers evaluated different prediction models across 17 diseases using these additional omics parameters.

They also matched important molecular attributes of the disease to demographic, clinical, and socioeconomic variables, illuminating underlying features across populations related to variability in disease etiology and risk.

The authors said, "Proteomics-only models generally outperformed metabolomics-only models for 16 of the 17 diseases, and integrating both omics added little prediction power over proteomics-only models."

Proteins steal the spotlight

Via this analysis, researchers identified key molecular markers such as the well-established prostate-specific antigen (PSA) in prostate cancer and new candidates such as PRG3 in skin cancer. Some of these proteins were stratified by medication use and also adjusted for socioeconomic status, demonstrating that omics data can correlate with individual risk factors, illuminate disease pathways, and potentially guide follow-up therapeutic targets.

The authors noted, "Our study not only develops better risk prediction models, but also provides insights into a wide range of disease risk factors."

The mixOmics tool was used to deal with data from extensive molecular datasets. Other deep learning methods, Cox regression, Elastic Net, Random Survival Forest, and MOGONET were also tested. Cox regression often overfitted; Random Forest models were slow; other approaches diluted the signal, but mixOmics was both accurate and fast.

Analysis of metabolites revealed only widespread lipid-related patterns and disease-specific structural changes associated with proteins. Collectively, these datasets revealed common biological pathways, metabolic processes, and signaling pathways involved in disease pathogenesis.

Proteins were the major predictors of combinations, confirming the strong individual predictive power of proteomics. It confirmed known protein-disease associations, such as PRG3 and skin cancer, and identified relationships between proteins and medications, demographics, and social factors.

Known drug-disease links, such as Bisoprolol and heart disease, were confirmed in some of the results, and others are promising leads for opportunities to repurpose drugs or identify new side effects that need further study.

Future directions for multi-omics

The study has several limitations. Hospital records may miss less severe cases that were managed in primary care settings. The analysis was performed on 159 metabolites and 2,923 proteins from plasma samples alone, failing to capture specific molecular events that may occur in internal organs.

All data were collected at baseline, thereby preventing assessment of disease progression over time. Only the two omics layers were the focus

of the study, and missing data have been imputed using methods that could introduce bias.

Feature importance could be misinterpreted due to correlations within the data. Since protein models proved so strong, the study focused on the best protein markers, without fully distinguishing which were common across diseases vs. disease-specific.

Despite limitations, the study carries important implications. Genomics and epigenomics enable greater predictive power than proteins and metabolites. Validating across populations enhances the robustness and fairness of a model. The combination of longitudinal data enables monitoring of disease progression, leading to timely interventions and personalized care processes.

More information: Jiawen Du et al, Multi-omics integration predicts the incidence of 17 diseases in the UK Biobank, *Nature Communications* (2026). DOI: 10.1038/s41467-026-73017-z

© 2026 Science X Network

Citation: Your blood may already know what illness comes next—long before symptoms appear, study finds (2026, May 14) retrieved 8 September 2026 from https://sciencex.com/news/2026-05-blood-illness-symptoms.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.