

ARTICLE | GUEST COMMENTARY

The missing links: how to ensure omics data fulfills its promise

Generalizable frameworks for longitudinal studies and interoperable data sharing systems needed to maximize omics impact

BY MATTHIAS EVERS, WERNER LANTHALER, KATARZYNA SMIETANA, TOBIAS SILBERZAHN AND MICHAEL STEINMANN

May 23, 2023 7:48 PM EDT



BioCentury & Getty Images

Omics data could herald a revolution in healthcare. The analysis of rich, interconnected and longitudinal multi-omics datasets promises a better understanding of the underlying biology of human health and disease, which in turn could lead to more effective prevention, earlier and more accurate diagnoses, new treatments and better choice of treatments.

Despite that promise, only a few countries have started to implement “omics” to benefit patient care and healthcare innovation. Why? Because it depends upon a seamless flow of omics data and insights running from R&D to patient care and back again.

Doing so will depend upon launching many more large, longitudinal omics studies that offer up rich data. But it will also depend upon overcoming the many barriers that can prevent omics data from being put to good use. Both issues are solvable.

Creating a framework for large, longitudinal omics studies that is patient-centric and generalizable would go a long way to help facilitate data generation, while sharing the data through interoperability would maximize its impact.

Creating a framework for large, longitudinal omics studies

Collecting longitudinal data at scale is a prerequisite for deriving high-quality omics insights. By tracking multiple types of omics over time in relation to an individual’s health, it is possible to move beyond a snapshot of any connections between them to a dynamic, systems model of the mechanisms that trigger the transition from health to disease and that drive disease progression. It may even become possible to identify mechanisms for preventing a disease from developing.

Since the Human Genome Project was completed, there have been various initiatives to create large repositories of omics and health data, including those by the Estonian Biobank, Genomics England and, in the United States, the National Institute of Health’s All of Us Research Program and Roadmap Epigenomics Mapping Consortium.

Even so, broad population studies generating rich omics data remain relatively rare, not least because of the challenge of ensuring enough participants enroll and then continue to engage for a sufficiently long period.

Key to establishing an effective framework for longitudinal studies is ensuring participants feel comfortable about data protection and that the study creates value for them or people like them.

According to Chris Wigley, the CEO of Genomics England, “For research participants to fully commit to long-term population studies, they need to be reassured their privacy and dignity will be protected. That’s table stakes. But it’s important to go beyond that, to ensure they are partners in the design and execution of such work.”

Listening to research participants and building in their thoughts and insights from their lived experiences “is not just the right thing to do — it also creates a much clearer path to impact for the research,” Wigley said.

People are particularly motivated to join a study when they identify with the study's goals.

While financial incentives might boost study numbers, and gamification, wearable technologies and various nudge interventions might help sustain engagement, it remains the case that people are particularly motivated to join a study when they identify with the study's goals and feel the knowledge gained could benefit them or people like them. This makes it particularly important to return results to individuals, even though it adds expense and complexity to the study, and to set compelling study goals that are broad enough to motivate considerable numbers to participate.

Robert Green, a professor of medical genetics at Harvard Medical School who has been a leader in studying the return of genomic information, notes that an overemphasis on privacy has consequences. "While millions of people have been part of genomics research studies, almost none of them who are carrying dangerous mutations for actionable conditions have learned about their risk status. This is an inexcusable result of the long-standing but inaccurate narratives in which medical science has over-indexed on distress and privacy and under-indexed on medical benefit."

Green recently co-led the first international recommendations on return of genomic results in research, which was published by the Global Alliance for Genomic Health. With regard to the broader topic of omics, he predicts: "Return of results will grow beyond just the genome. As specific information from multi-omics becomes clinically useful, we should take steps to return these results as well."

But crafting an effective longitudinal study is not just about returning results, or motivating people to join and stick around; it's about motivating the right people. Studies could deliver insights more rapidly by focusing initial enrollment on participants at high risk of developing the disease of interest — diabetes or cardiovascular diseases, for example — due to genetics or other factors, or by focusing on diseases that exhibit a "flare and relapse" pathology, such as psoriasis or cluster headaches.

It is imperative that any framework for a longitudinal omics study include guidelines for ensuring the participants enrolled reflect the diversity of the patient populations with the diseases of interest.

Building bridges between data generation and data impact

Vast amounts of data are already available that could prove beneficial to researchers, healthcare providers and patients. But the fact that they derive from so many different sources and that there are no standards or protocols for collecting the data make it hard to use. Another reason the data are underutilized — and will continue to be so — is that it is not routinely shared.

To overcome this, bridging mechanisms are needed to make sure the data generated has impact. We see three.

The first is guidelines and/or legal frameworks that protect privacy but still allow people to contribute their medical data to research. Existing guidelines don't always strike the right balance.

The second is a certification process that can vouch for the robustness of published data and a study's adherence to a defined protocol. It could, for example, certify that the data was generated using a certain set of standards, common ontologies, protocols for conducting and recording observations (e.g., ensuring machine-readability), and mechanisms for systematic and consistent follow-ups and linkages to other medical datasets.

A certification process could address some of the concerns about the quality and actionability of the data and promote interoperability. As it stands, some databases are not interoperable even within the same country.

To help users of the data quickly assess whether it meets their needs, the portal could label and describe the data in terms of its sources, breadth, depth and longitudinality, for example.

Sharing must become a routine part of R&D processes if the full value of omics data is to be realized.

The third bridge needed is to promote data sharing. Even if new datasets are built, privacy concerns are addressed, and steps are taken to verify the robustness of the data, mechanisms might still be needed to encourage more of that data to be shared.

Sharing must become a routine part of R&D processes if the full value of omics data is to be realized.

This is particularly important in early research, where omics data could help researchers predict which targets and molecules might be successful in clinical trials, thereby reducing the number of late-stage drug development failures and risks to trial participants, speeding the development of successful new therapies, and improving the understanding of which patient populations could benefit from scientific breakthroughs. In short, sharing more data supports more effective and safer drug development.

Werner Lanthaler, CEO of Evotec AG (Xetra:EVT), predicts fundamental changes to the drug discovery process in the near future based on use of omics data to help predict which targets and molecules will be successful: "Pan-omics driven drug discovery will completely change current workflows. We will only start to significantly invest in drug targets once we have a much better disease understanding, and once we can predict true patient relevance and safety. With this, the 'science fiction' of drug discovery from a decade ago will become reality."

Working practices would likely have to change, however. All members of the medical ecosystem — physicians in the real world, clinicians conducting trials, biostatisticians, data scientists and bench researchers — would have to become comfortable sharing far more than just a handful of key endpoints and positive results, as is the norm today. Of equal or even more interest would be data on failed experiments and broader sets of exploratory variables, such as granular phenotyping information and patient-reported outcomes. The data generated by innovative clinicians could, for example, be used by both researchers exploring new treatment hypotheses and data scientists developing predictive models describing the behavior of a disease.

Peer pressure might help encourage both the building and sharing of such datasets but will unlikely suffice. Mandates could prove more effective — perhaps one that requires all those who receive public funding or grants from foundations to publish data from their experiments, whether successful or not. Sharing could also be encouraged by asking research centers to state the proportion of data points made available to the broader community that conformed to agreed standards — and to take responsibility for the validity of the data submitted, including the timely retraction of results that proved to be false or not replicable.

Getting started

We do not underestimate the difficulty of implementing any of this. A concerted push by major stakeholders would be required.

Efforts by influential organizations, an intergovernmental group, bold philanthropists or entrepreneurs could help set up longitudinal studies that would act as catalysts, spurring others to follow suit.

Collaboration across the entire industry would likely be needed to introduce a system that certified the robustness of published data and a study's adherence to a defined protocol, while public-private consortia might be well-positioned to begin building the bridges required to ensure omics data is shared. The industry has a good track record in areas such as preclinical toxicology and clinical trial conduct, where several shared standards were established, and data sharing efforts led to tangible benefits. Moreover, the industry's voice would be key in discussions on issues such as whether data-sharing could be considered pre-competitive.

Despite the challenges, the prize surely merits efforts to overcome them. The facilitating of large-scale omics data studies and information flows between clinicians, health systems, and researchers would enable the development of an interconnected map of human health and disease that drives better decisions, both at the bench and the bedside.

Matthias Evers is Chief Business Officer and Werner Lanthaler is CEO of Evotec. Katarzyna Smietana is an associate partner in McKinsey's Wroclaw office, Tobias Silberzahn is a partner in McKinsey's Berlin office and Michael Steinmann is a managing partner in McKinsey's Zurich office and leads McKinsey's Life Sciences R&D Practice.

This article represents some of the views expressed at Monastery & Minds 2022, an event attended by some 30 executives, investors and other members of the life sciences and healthcare sectors. Two BioCentury editors participated in the meeting.

Purchase Reprint/Distribution Rights

This article may not be distributed to non-subscribers.

Purchase this article for limited one-time distribution and website posting.

[Purchase Rights](#)

BCIQ Company Profiles

[Evotec SE](#)



About BioCentury