The big medical data miss: challenges in establishing an open medical resource

Eric J. Topol

I call for an international open medical resource to provide a database for every individual's genomic, metabolomic, microbiomic, epigenomic and clinical information. This resource is needed in order to facilitate genetic diagnoses and transform medical care.

> "We are each, in effect, one-person clinical trials"

Laurie Becklund was a noted journalist who died in February 2015 at age 66 from breast cancer. Soon thereafter, the *Los Angeles Times* published her op-ed entitled "As I lay dying" (REF. 1). She lamented, "We are each, in effect, one-person clinical trials. Yet the knowledge generated from those trials will die with us because there is no comprehensive database of metastatic breast cancer patients, their characteristics and what treatments did and didn't help them". She went on to assert that, in the era of big data, the lack of such a resource is "criminal", and she is absolutely right.

Internet of DNA

Around the same time of this important op-ed, the *MIT Technology Review* published their issue entitled "10 Breakthrough Technologies 2015" and on the list was the "Internet of DNA" (REF. 2). While we are often reminded that the world we live in is becoming the "Internet of Things", I have not seen this terminology applied to DNA before. The article on the "Internet of DNA" decried, "the unfolding calamity in genomics is that a great deal of life-saving information, though already collected, is inaccessible". It called for a global network of millions of genomes and cited the Matchmaker Exchange as a frontrunner. For this international initiative, a growing number of research and clinical teams have come together to pool and exchange phenotypic and genotypic data for individual patients with rare disorders, in order to share this information and assist in the molecular diagnosis of individuals with rare diseases.

Global medical resource

These callouts for a global resource for cancer and rare-disease genomics are just the start. Why don't we have a global medical resource of every individual's geographic information system (GIS) that includes not only demographics, phenotype and DNA sequence, but also their physiome and exposome (via sensors), anatome (via medical scans) and the rest of their biological data (transcriptome, metabolome, proteome, microbiome and epigenome)[3]? This multilayered medical map of each individual could be used in conjunction with each person's treatment and outcomes to inform the care of future patients with similar presentations. Once such a resource is established, the next person who comes along with a similar disorder, or in quest of preventing a serious condition, would have the benefit of the best information possible — not from the literature or animal models but from very closely resembling *Homo sapiens*.

For example, in a project that we have undertaken at the Scripps Translational Science Institute to understand the molecular basis of sudden death in victims with a negative autopsy, we sequenced the DNA from deceased and family members. Even after extensive biocomputing and filtering, the main result is typically a long list of variants of unknown significance (VUS). To overcome this challenge, an Internet of DNA — or what I have referred to as a massive, open, online medicine resource (MOOM) — would help to quickly identify the genetic cause of the disorder[4] and, in the process of doing so, precious guidance for prevention, if necessary, would become available for such families who are currently left in the lurch as to their risk of suddenly dying.

So why aren't such MOOMs being assembled? By 2020, it has been projected that ~5 million individuals will have undergone whole-genome sequencing, producing hundreds of petabytes of data[5]. But our digital infrastructure, with seemingly limitless cloud and super computing, can handle this scale of data with ease and, at the same time, ingest the exabytes generated from real-time biosensor streaming of data. The start-up Truecaller is well on its way to establish a mobile app

Scripps Translational Science Institute, 3344 N. Torrey Pines Court, La Jolla, California 92037, USA.
e-mail: etopol@scripps.edu
doi:10.1038/nrg3943

> “a massive, open, online medicine resource … would help to quickly identify the genetic cause of the disorder”

to access every phone number on the planet. So capturing or storing data isn't an obstacle to establishing this scale of a medical database. While analysing and extracting all the meaningful information from large genomic data sets is currently a bottleneck, this is not an insolvable problem.

There has also been much discussion related to privacy concerns that patients might be unwilling to participate in a massive medical information resource. However, multiple global consumer surveys have shown that more than 80% of individuals are ready to share their medical data provided that they are anonymized and their privacy maximally assured[4]. Indeed, just 24 hours into Apple's ResearchKit initiative, a smartphone-based medical research programme, there were tens of thousand of patients with Parkinson disease, asthma or heart disease who had signed on. Some individuals are even willing to be “open source” — that is, to make their genetic and clinical data fully available with free access online, without any assurance of privacy. This willingness is seen by the participants in the recently launched Open Humans initiative. Along with the Personal Genome Project, Go Viral and American Gut have joined in this initiative. Still, studies suggest that most individuals would only agree to be medical research participants if their identities would not be attainable. Unfortunately, to date, little has been done to protect individual medical privacy, for which there are both promising new data protection technological approaches[4] and the need for additional governmental legislation.

This leaves us with perhaps the major obstacle that is holding back the development of MOOMs — researchers. Even with big, team science research projects culling together hundreds of investigators and institutions throughout the world, such as the Global Alliance for Genomics and Health (GA4GH), the data obtained clinically are just as Laurie Becklund asserted in her op-ed — “one-person clinical trials” (REF. 1). While undertaking the construction of a MOOM is a huge endeavour, there is little motivation for researchers to take on this task, as this currently offers no academic credit and has no funding source. But the transformative potential of MOOMs to improve medical care is extraordinary. Rather than having the knowledge die with each of us, the time has come to take down the walls of academic medical centres and health-care systems around the world, and create a global knowledge medical resource that leverages each individual's information to help one another.

1. Becklund, L. As I lay dying. *Los Angeles Times* [online], http://www.latimes.com/opinion/op-ed/la-oe-becklund-breast-cancer-komen-20150222-story.html#page=1 (2015).
2. Ten Breakthrough Technologies 2015. *MIT Technology Review* [online], http://www.technologyreview.com/lists/technologies/2015/ (2015).
3. Topol, E. J. Individualized medicine from prewomb to tomb. *Cell* **157**, 241–253 (2014).
4. Topol, E. J. *The Patient Will See You Now* (Basic Books, 2015).
5. Topol, E. J., Steinhubl, S. R. & Torkamani, A. Digital medical tools and sensors. *JAMA* **313**, 353–354 (2015).

Competing interests statement
The authors declare competing interests: see Web version for details.

FURTHER INFORMATION
American Gut: http://americangut.org/
Apple's ResearchKit: https://www.apple.com/researchkit/
Global Alliance for Genomics and Health: http://www.genomicsandhealth.org
GoViral: https://www.goviralstudy.com/
Matchmaker Exchange: http://www.matchmakerexchange.org
Open Humans: https://www.openhumans.org
Personal Genome Project: http://www.personalgenomes.org/
Truecaller: https://www.truecaller.com

ALL LINKS ARE ACTIVE IN THE ONLINE PDF