# Rich data sets could end costly drug discovery

# Deep cohorts for drug discovery

## (Rich data sets could end costly drug discovery)

E. Segal

How deep phenotyping of human cohorts may revolutionize drug development

By charging $2.1 million dollars per patient for its new spinal muscular atrophy (SMA) treatment, Novartis recently broke the world record for the most expensive drug[1]. But with so many compounds failing and at a cost of $3-10 billion dollars to develop a new molecular entity into a therapy[2], perhaps it's no wonder that drugs have become so expensive. Can we do better?

Yes, we can. One way is by using the vast amount of genetics data collected on many human individuals, since drugs with genetic evidence linking their target to the disease are twice as likely to complete development and be approved[3,4]. Genetics data helps because unlike traditional drug development that integrates human data only after other cellular and pre-clinical testing (typically years into the process), genetics provides evidence that the drug target is relevant to the disease right from the start, and in human subjects.

Following the same logic, we can optimize drug discovery much more if we start with even deeper phenotyping of human subjects using the most advanced multi-omics, body sensors, and imaging technologies. The idea of using humans as a model for humans is not new[5], but the ability to profile many individuals with modern multi-omics methods is new and as a result, more and more such initiatives are underway[6-9] with the goal of analyzing the collected data for new diagnostics and therapeutics.

Take the microbiota, for example. This vast collection of trillions of bacteria that reside throughout our body has been implicated in conditions ranging from infections and autoinflammatory disease to autism, cardiovascular disease, and oncology[10]. Growing evidence, primarily from animal models[11] (but also from humans[12]), suggests a causal effect for the microbiota in disease, e.g., by regulating host gene expression[13] or by being involved in the metabolism of metabolites that circulate in our blood[14,15]. A notable example is trimethylamine N-oxide (TMAO), which is derived from gut microbial metabolism of choline and carnitine, and is a marker for cardiovascular disease[16]. Microbiota-derived metabolites that reach the millions of neurons in the intestine that connect to the brain may also pass the blood–brain barrier. The microbiota can also affect the efficacy and toxicity of pharmaceutical drugs, as in the case of a bacterial enzyme found to metabolize the Parkinson's drug L-dopa[17], and in the response to cancer immunotherapy[18].

But beyond its potential role in driving disease, an appealing feature of the microbiota is that it is predominantly shaped by modifiable environmental factors such as diet, and not by genetics[19].

Thus, devising dietary interventions that target bacteria that synthesize TMAO may help lower cardiovascular disease risk (if TMAO proves to be causative). Similarly, dietary changes that target bacteria that interfere with drug metabolism may be an effective companion treatment for cancer immunotherapy. And the regulatory path for approving such 'microbiome nutrition[20]' interventions is much easier.

Developing such microbiota-related therapies requires two components. First, a causal effect in disease must be established for some microbial target. Second, means of modifying this target must be found. Data from carefully designed human cohorts can guide the discovery of both components. Putative targets may first be proposed by finding associations between microbiota composition and disease markers, followed by methods for establishing causality based on human genetics data or animal models. Interventions that modify bacterial targets may be found by collecting detailed dietary and microbiota data from many individuals, deriving models of how dietary intake affects microbiota composition, and validating the models with controlled dietary interventions. Starting both steps with human data ensures that we pursue human-relevant bacterial targets. In contrast, targets with therapeutic value in animal disease models may not work in humans, either because they are less abundant or act through different mechanisms.

It is still early days for microbiota-related therapeutics and many fundamental challenges remain, including establishing causal mechanisms and understanding how best to manipulate our microbiota (e.g., duration, frequency). However, with over 5 million bacterial genes, the microbiota at least has the potential to be a prolific reservoir of modifiable targets with therapeutic effects.

When the causal mechanism is unknown, we may still be able to use human cohorts with microbiome data to devise new therapies, by utilizing predictive algorithms and subsequent testing in randomized clinical trials (RCTs). For example, we previously sought to target blood glucose levels after meals as these are important for several indications, including obesity and diabetes management and prevention. We started directly with a cohort of 900 human subjects in which we continuously tracked blood glucose and obtained microbiota, genetics, metabolomics, blood tests, diet, lifestyle, and physiological measurements[21]. Our data of ~50,000 blood glucose responses to meals revealed that different people have vastly different responses to the same meal, and we devised a machine learning algorithm that accurately predicted these personalized blood glucose responses from clinical and microbiome data. We further showed in both short-term[21] and 12-months[22] RCTs, that personalized dietary interventions based on the algorithm successfully balanced glucose levels in people with prediabetes, significantly outperforming the standard-of-care diet. Thus, for dietary interventions discovered by human data analysis, therapeutic potential can be evaluated by going directly to human RCTs.

We greatly need new approaches to drug development. Human multi-omics and physiological measurements at scale, or 'deep cohorts', offer one way to prioritize human-relevant targets for further development and eventual testing in RCTs. However, to be useful, careful considerations should be given to their construction. First, the type and depth of the data measured should be relevant to the disease studied and in the right scale (e.g., molecular, organ, or whole-body). Having all types of data on the same cohort can then be powerful for comparing their relative

contributions and for getting us closer to causal candidates by identifying the data type with the strongest association. Deep profiling also allows us to define more exact targets (e.g., using continuous glucose monitoring vs. a one-time fasting glucose measure) and to discover novel disease biomarkers and drug targets (e.g., microbiome-based). It also allows modeling disease state or treatment response as a continuum and avoid dichotomies and arbitrary disease thresholds[23]. This may lead to better estimates of patients' disease risk and better prioritization of patients to treatments and RCTs.

Obtaining longitudinal measurements of the same subjects over time is another critical component to integrate into deep cohorts, since such N-of-1 studies[23,24] bypass confounders of interpersonal variability by having subjects serve as their own control. Finally, since resources are always limited, cohort size is another important consideration. Thousands or tens of thousands of subjects strike a good balance that still allows longitudinal, deep multi-omics profiling.

The time may be right to use deep cohorts to transform drug discovery from its laborious, costly, risky, and time-consuming era. We simply can't afford another world record.

**References**


1. Stein, R. At $2.1 Million, New Gene Therapy Is The Most Expensive Drug Ever (https://tinyurl.com/yxge9v7s). (2019).
2. Schuhmacher, A., Gassmann, O. & Hinder, M. Changing R&D models in research-based pharmaceutical companies. *J Transl Med* **14**, 105 (2016).
3. Nelson, M.R. *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856-60 (2015).
4. King, E.A., Davis, J.W. & Degner, J.F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *bioRxiv* **doi.org/10.1101/513945**.
5. FitzGerald, G. *et al.* The future of humans as model organisms. *Science* **361**, 552-553 (2018).
6. Gaziano, J.M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* **70**, 214-23 (2016).
7. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
8. Chen, Z. *et al.* China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* **40**, 1652-66 (2011).
9. All of Us Research Program, I. *et al.* The "All of Us" Research Program. *N Engl J Med* **381**, 668-676 (2019).
10. Cully, M. Microbiome therapeutics go small molecule. *Nat Rev Drug Discov* **18**, 569-572 (2019).
11. Ridaura, V.K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214 (2013).
12. Sanna, S. *et al.* Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat Genet* **51**, 600-605 (2019).
13. Richards, A.L. *et al.* Gut Microbiota Has a Widespread and Modifiable Effect on Host Gene Regulation. *mSystems* **4**(2019).
14. Fischbach, M.A. Microbiome: Focus on Causation and Mechanism. *Cell* **174**, 785-790 (2018).
15. Wikoff, W.R. *et al.* Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A* **106**, 3698-703 (2009).
16. Li, X.S. *et al.* Gut microbiota-dependent trimethylamine N-oxide in acute coronary syndromes: a prognostic marker for incident cardiovascular events beyond traditional risk factors. *Eur Heart J* **38**, 814-824 (2017).
17. Maini Rekdal, V., Bess, E.N., Bisanz, J.E., Turnbaugh, P.J. & Balskus, E.P. Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science* **364**(2019).
18. Routy, B. *et al.* Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **359**, 91-97 (2018).
19. Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210-215 (2018).
20. Lam, Y.Y., Zhang, C. & Zhao, L. Causality in dietary interventions-building a case for gut microbiota. *Genome Med* **10**, 62 (2018).
21. Zeevi, D. *et al.* Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **163**, 1079-94 (2015).
22. Personalized Nutrition for Pre-Diabetes (https://ClinicalTrials.gov/show/NCT03222791). (2016).
23. Senn, S. Statistical pitfalls of personalized medicine. *Nature* **563**, 619-621 (2018).
24. Schork, N.J. Personalized medicine: Time for one-person trials. *Nature* **520**, 609-11 (2015).