ORIGINAL ARTICLE

Real-Time Assessment of Wellness and Disease in Daily Life

Dennis Ausiello^{1,*} and Scott Lipnick¹

Abstract

The next frontier in medicine involves better quantifying human traits, known as “phenotypes.” Biological markers have been directly associated with disease risks, but poor measurement of behaviors such as diet and exercise limits our understanding of preventive measures. By joining together an uncommonly wide range of disciplines and expertise, the Kavli HUMAN Project will advance measurement of behavioral phenotypes, as well as environmental factors that impact behavior. By following the same individuals over time, KHP will liberate new understanding of dynamic links between behavioral phenotypes, disease, and the broader environment. As KHP advances understanding of the bio-behavioral complex, it will seed new approaches to the diagnosis, prevention, and treatment of human disease.

Key words: big data analytics; big data architecture; big data industry standards

Introduction

The field of genetics has exploded since the mapping of the human genome, but despite the treasure trove of information it provides, the human genome alone does not liberate comprehensive understanding of the human condition. The next great leap in human measurement and analysis is mapping the phenome: the sum total of human traits. The Kavli Human Project (KHP) will greatly advance this process.

Advances in human genetics have identified many genetic contributors to disease risk. At the same time, there are ever more extremely large datasets available containing multiple types of data relevant to human health. Growing analytic and computing capabilities are enabling mining of these large datasets for insights at the level of individual patients or entire populations. Yet the methods used to diagnose disease have conspicuously lagged behind these recent exciting discoveries. This is because disease traits are also influenced by largely unmeasured environmental and behavioral factors.

The KHP will liberate progress by convening a wider range of expertise than is traditional, including device engineers, front-line physicians, geneticists, and experts in behavioral modeling. It will focus directly on the development of novel quantitative human measurements, or phenotypes. By tracking individuals in rich detail over a long period of time, the KHP will capture and catalog dynamic patterns of individual and social behavior in new and richer metrics than ever before. This will help us develop new approaches to measuring human health, and so we can quantify wellness and disease in a more continuous manner, rather than in the current episodic manner. It will allow us to combine data on individual genetics with new human phenotypes at multiple levels, including functional characterization of patient-derived cells, specific physiologic pathways, diet, the microbiome, and wearable physiologic sensors. With its fixed geographic base in New York, the KHP will enable us to measure environmental exposures, including potentially inhaled or ingested toxins, which are currently poorly

¹Center for Assessment Technology and Continuous Health, Massachusetts General Hospital, Boston, Massachusetts.

Some of the material presented in this article has been previously published (Ausiello D, Shaw S. Quantitative human phenotyping: The next frontier in medicine. *Trans Am Clin Climatol Assoc.* 2014;125:219–226; discussion 226–228) and is used with permission of the author, publisher, and The American Clinical and Climatological Association.

*Address correspondence to: Dennis Ausiello, Department of Medicine, 55 Fruit Street, Boston, MA 02114, Email: Ausiello.Dennis@mgh.harvard.edu

measured. This will create another important new data resource that can be integrated with genotypic, phenotypic, and clinical information.

Over time, integrating phenotypes at multiple scales of biology (from cells to the whole individual) will likely become the new norm in human disease studies. This will benefit human health in numerous ways, from individual patient empowerment to biomedical discovery, as well as new approaches to diagnosis, prevention, and precision therapeutics.

In the remainder of the article we first discuss key advances in biological understanding and the failure as yet to realize the full potential of these discoveries to improve human health. We then sketch out the need for new behavioral measurements to advance understanding of the bio-behavioral complex followed by a discussion of how the design of the KHP can liberate such advances in measurement. We conclude by describing how the first implementation of such an endeavor—the newly formed Center for Assessment Technology and Continuous Health (CATCH) at Massachusetts General Hospital (MGH)—illustrates the high potential that increased monitoring has for improving health outcomes.

Human Biology, Genetics, and Health

Advances in human genetics (especially genome-wide association studies) have identified unprecedented numbers of chromosomal loci that contribute to human traits and risk of disease.¹ These genetic data, combined with insights from basic, hypothesis-driven laboratory research, provide a much clearer outline of the genes that contribute to disease (the parts list), including many that were previously unsuspected. The discovery of so many disease loci promises to remake our understanding of disease mechanisms and susceptibility. Particularly exciting is the prospect that new biomarkers of disease risk can be identified.

While the mapping of the genome is the most widely recognized advance in our understanding of human biology, it is but one piece of the human puzzle—there are many other ongoing advances happening, as well. One key scientific advance lies in the emergence of the human microbiota as an important contributor to many chronic diseases. The community of approximately 10^{14} bacterial, archaeal, fungal, and viral cells or particles that reside on each individual constitutes the human microbiota; the “microbiome” additionally refers to the genetic materials and product biomolecules of the microbiota. Microbes colonize the gut at birth

and the community is shaped by diet, hygiene, infections, drugs, other environmental exposures, and host genetics. Recent studies in normal volunteers and disease cohorts (such as those with inflammatory bowel disease, obesity, diabetes, or cardiovascular disease) are revealing how all of these genetic and environmental factors can shape what types of microbes are present, and their aggregate influence on human metabolism and immunity (and even behavior in some animal models). (See refs.^{2–8} for more study details.)

A third set of advances relates to our understanding of biological pathways. For example, a wide variety of inflammatory cells and pathways are being studied in auto-inflammatory disease as well as common diseases such as type 2 diabetes and cardiovascular disease. Biomarkers such as C-reactive protein or the erythrocyte sedimentation rate are nonspecific markers of inflammation currently used in clinical practice. A variety of new approaches have the potential to enable scientists to parse inflammation more precisely, including serum levels of specific cytokines or mediators, or assays of the activity of inflammatory cells (including molecular imaging of inflammatory cells, or microfluidic devices that can trap or analyze single cells).

In sharp contrast to the ongoing revolution in biological understanding, translation into medical practice has been taking place at a far slower pace. This is particularly true for the clinical translation of genetic discoveries into new approaches to diagnosis, prevention, and treatment. For the vast majority, there has been little to no improvement in treatment or in health. Consider the cases of type 2 diabetes and lipid disorders, each of which is now associated with many dozens of chromosomal regions that influence disease risk. Despite this, these conditions are largely monitored using blood tests that have been used for decades: glycosylated hemoglobin and blood glucose for diabetes, and a lipid panel consisting of fasting total cholesterol, low-density lipoprotein, high-density lipoprotein, and triglycerides.

In another revealing example, inclusion of a genetic risk score (based on several validated variants from genome-wide association studies) failed to improve the ability to predict 10-year risk of developing cardiovascular disease compared to traditional assessments (including factors such as age, gender, smoking history, and the presence of diabetes, hypercholesterolemia, hypertension, or a family history of cardiovascular disease).⁹ Despite the revolutions that are taking place in

our biological understanding, the available methods used to diagnose and quantitate disease have conspicuously lagged.

More broadly, while the large number of chromosomal loci newly implicated in many diseases represents a true scientific *tour de force* with tremendous future potential applications in medicine, it remains a challenge to effectively use genetic information to stratify risk. This is likely due to several factors, including the sheer number of genetic loci that can contribute to individual risk and, perhaps most critically, the importance of largely unmeasured environmental and behavioral factors that influence risk of important conditions such as obesity, type 2 diabetes, cardiovascular disease, and cancer. The comprehensive approach to information gathering to be employed by the KHP offers the opportunity to overcome this barrier.

The Bio-Behavioral Complex

A key problem holding back medical advances is that any disease trait is under the influence of not just many genetic loci, but also environmental and behavioral influences. Our understanding of and ability to measure behavior has not advanced at anywhere near the same pace as our ability to understand and measure biological factors. Hence, there is a fundamental need for new approaches to measure human health to better quantify wellness and disease in a more continuous manner as our patients lead their daily lives.

The high potential of the KHP lies precisely in its focus on measuring not just biology but also behavior and the interactions between them. Just as genotyping and genetic sequencing technology have progressed rapidly, an analogous renaissance for phenotypes is required to enable human measurements with greater physiologic resolution and lower cost. For instance, the process of deciphering the physiologic and health consequences of disease-related genetic risk alleles is laborious, expensive, and largely limited to existing phenotypes investigated in small clinical studies. Similarly, novel therapeutic agents are being developed with unprecedented mechanisms but are often evaluated using outdated phenotypes. Novel phenotypes are needed that are more specific and proximate to the mechanisms being modulated. This will enable more rapid testing of novel therapeutic hypotheses in humans, and thus earlier views on the potential efficacy of new agents. Therapeutic trials would also benefit from better stratification of patient subsets to enrich trial populations for those most likely to respond. Stratification

by specific genetic mutations has enabled dramatic progress in targeted therapies in certain types of malignancies, such as nonsmall cell lung cancer bearing mutations in the EGF receptor or the ALK kinase. However, for the majority of genetically complex and chronic diseases that are not driven by somatic mutation (as these specific cancer subtypes appear to be), the optimal stratification is likely to come from a combination of genetic and phenotypic stratification.

A new approach to human measurements can also transform how individuals engage in their own health, provide insightful measurements in real time, and allow individuals to monitor and improve their own health and wellness (in partnership with their physicians and caregivers). A key challenge is to move the monitoring of health and disease away from the physical and time constraints of physician offices and hospitals, and into the domain of patients' lives. The ability to track symptoms or health status more quantitatively can help patients and their caregivers understand disease trends and how interventions may worsen or ameliorate symptoms, and allow the time during an office visit to be used more effectively.

Another group of poorly measured factors relates to diet and environmental exposures, which may potentially contribute inhaled or ingested toxins. Exposures are typically accessed on rare occasions through survey instruments based on recall, blood, or urine assays. While continuous measurement of environmental exposures may not be necessary (or feasible), enabling more facile and passive quantification of environmental exposures will create an important new data resource that can be integrated with genetic and clinical information.

Continuous Measurement, Big Data, and the KHP

Our current system of delivering healthcare is episodic and reactive. That is, patients see their physicians largely at regularly scheduled intervals (typically one year), and/or when symptoms appear or worsen. At a time when the healthcare system in the United States faces tremendous pressure to contain costs and improve efficiency and outcomes, this episodic approach forces patients to summarize and communicate months of symptoms and observations in a brief office visit, and limits the ability of patients and physicians to proactively address emerging medical issues.

During their episodic appointments, the methods physicians use to assess disease in our patients have largely remained the same for decades. The typical

office visit will document the patient's medical history and symptoms since the last visit (usually several months ago); parameters such as weight, heart rate, blood pressure, and respiratory rate; a physical examination; and perhaps standard blood tests such as general chemistry values and a lipid panel. While specialized blood diagnostics and imaging studies are used to investigate specific diagnoses, the most commonly used measures reflect an uneasy balance between cost, the time constraints of an office visit, and the ability to detect significant changes in health status.

We are fortunate to live in an era in which increased behavioral and biological measurement is technologically possible. The increasing availability of digital and genetic data, and measurement platforms thereof, facilitates the collection and analysis of extremely large datasets containing multiple types of data relevant to human health. These include the data contained in electronic medical records (EMRs), genetic data (e.g., characterization of mutations in a tumor biopsy, or genome-wide genotyping), and pharmacy or claims data. Less traditional but growing sources of data include personal fitness trackers (such as wearable activity monitors), online social communities and other web forums, and medical measurements such as blood pressure or blood glucose that can be transmitted wirelessly. Growing analytic and computing capabilities are enabling mining of these large datasets for insights at the level of individual patients or entire populations, and can have a profound impact on how healthcare is administered in the United States in the future.

Certain types of medical data are already collected continuously, and represent opportunities for data repurposing. For instance, millions of implanted cardiac devices such as pacemakers and implantable cardioverter-defibrillators provide ready access to beat-by-beat heart rate data. Continuous glucose monitors (typically accessed via a small sensor in the interstitial space) have long been used primarily to guide dosing of automated insulin infusion pumps, but may yield insights into the dynamics of glucose regulation.

Other important behaviors to measure include exercise, diet, and medication adherence and make significant contributions to several diseases ranging from cancer to diabetes and cardiovascular disease. Data from wearable devices such as activity monitors (such as digital pedometers) and wrist-based monitors (e.g., devices that measure skin galvanic response as a reflection of stress) could provide insight into individual behaviors as well as facilitate feedback. Several types of

wearable measurements represent physiologic parameters that could be analyzed in certain disease-specific contexts, but also represent important sources of information for individuals as they monitor their own health.

Studies suggest that individuals commonly discontinue wearable devices after several months, potentially limiting their widespread application. But embedding sensors into devices that are ubiquitous and used with high persistence, such as mobile phones, may open up new avenues. Furthermore, thanks to increased capabilities, mobile phones are increasingly used for a variety of routine behaviors such as communication, travel location, and even specific health-related software ("apps"), and the mobile phone represents an appealing platform for a variety of continuous and behavioral measurements.

But the biggest opportunity lies in the fact that we are currently in the midst of unprecedented technological change. A profusion of powerful computing and communications platforms has been enabled by small form factors and ubiquitous Internet access. These include smart phones or tablet computers with cellular network and/or wireless Internet access, and personal devices that can transmit digital health-related information. These devices and associated software or apps place unprecedented data collection, retrieval, and exchange capabilities literally in the hands of individuals, and liberate these activities from traditional location-based constraints (such as physician offices or hospitals). However, these powerful capabilities are only beginning to be systematically explored in the context of individual health.

The importance of the ongoing mobile revolution is that it makes the gathering of measurements more unobtrusive to the individual. Intermediate benefits may be realized by current "mobile health" efforts that use traditional measurements such as blood glucose or weight and simply transmit them to caregivers (e.g., through an iPhone attachment that measures blood glucose or wireless-equipped weight scales). But a full realization will require quantitative measurements that can be collected passively. This has spawned great interest in the so-called wearable sensors, such as devices worn on the waist or wrist, embedded in smart phones, or even embedded in clothing that could reflect physiologic parameters such as heart rate and respiration, activity, stress, or behavior. Passive data collection significantly increases the completeness of data captured about the populations to which this approach may be applied; active data entry risks biasing the study population toward

participants who have higher levels of technological familiarity or higher degrees of motivation or engagement in their health. More complete, less biased datasets will better allow analysis to reveal the effect of therapeutic or other interventions.

The KHP will liberate progress by convening a wider range of expertise than is traditional, including device engineers, front-line physicians, geneticists, and experts in behavioral modeling focused directly on the development of novel phenotypes. By tracking individuals in rich detail over a long period of time, the KHP will capture and catalog dynamic patterns of individual and social behavior in new and richer metrics than ever before. This will help us develop new approaches to measuring human health, and so we can quantify wellness and disease in a more continuous manner, rather than in the current episodic manner.

With its fixed geographic base in New York, the KHP will enable us to measure environmental exposures, including potentially inhaled or ingested toxins, which are currently poorly measured, providing consistency across the study population that could not be achieved with a more disaggregated geographic study frame. This will create another important new data resource that can be integrated with genotypic, phenotypic, and clinical information.

Given the multiple modalities of measurement that it is designed to include, the KHP will liberate studies that combine data on individual genetics with new human phenotypes at multiple levels, including functional characterization of patient-derived cells, specific physiologic pathways, diet, the microbiome, and wearable physiologic sensors. Because of the diversity of this universe of potential measurements, the development of integrative analytics that allow disparate data types (traditional and nontraditional) to be analyzed in concert will be critical. Integrating phenotypes at multiple scales of biology (from cells to the whole individual) will likely become the new norm in human disease studies.

Continuous Monitoring, CATCH, and the Future of Healthcare

An ongoing effort taking place at MGH illustrates the high potential that increased monitoring has to change the current episodic healthcare paradigm and improve health outcomes. The newly formed CATCH seeks to discover and apply new ways to quantitatively measure human phenotypes in health and disease. Through a multidisciplinary collaboration of scientists, physicians,

engineers, computer scientists, and behavioral experts across MGH, the Massachusetts Institute of Technology, and the private sector, CATCH will leverage the digital and genetic revolutions to transform how individuals monitor their own health, and how physicians can prevent, diagnose, and treat disease.

To fully implement this vision, important changes are also needed in the culture of patient care and scientific research. Scientific collaborations will need to convene a wider range of expertise than is traditionally sought, including device engineers, front-line physicians, geneticists, and experts in sociology and behavior. Collection of these novel data types will require new approaches to data ownership and security that appropriately balance an individual's control over use of their data and with permission and trust framework for secondary use of data in specific contexts. Analyses must be focused on actionable insights and rendered visually to allow patients and their caregivers to understand the medical implications.

CATCH provides an important model for the collection of comprehensive phenotypic data, demonstrating the promise of this approach. The KHP will capitalize and expand upon the ideas and lessons coming out of CATCH in order to successfully characterize the bio-behavioral complex on a large scale. Together, these two projects will enhance our understanding of the complex forces that shape human health.

Implementation in the KHP

Investigations that are outlined in this article would utilize the following KHP data sets, among others:

1. Medical information on study participants' health would be available from the medical history and records going forward (EMRs, doctor's notes, hospital records, dental records). Prescription data would be gathered via the NY State Prescription database. This information would be complemented by the SPARCS database and KHP's own tests: blood tests (blood metals, vitamins, lipids, glucose, and other biomarkers), and urine and hair tests (smoking, alcohol, and substance use) every three years.
2. Information on genetics would be gathered via whole genome sequencing of blood samples for adults (saliva for children) performed at study intake. In addition, data on epigenetics would be gathered via triennially performed assays.
3. Microbiome (oral and gut) data would be gathered via periodic saliva and stool samples. In addition,

these samples would be targeted for collection during each health or emotional crisis.

4. Dietary data would be collected via periodic food diaries, complemented by mining for food purchases in financial data (credit cards, debit cards, checks).
5. Mobility and activity level information would be collected by activity trackers and smartphone apps.
6. Information on financial status and participation in government assistance programs (SNAP, Social Security, TENF) would be available via financial data gathered using a combination of automated and survey-based methodologies.
7. Connectivity to existing and future medical devices would be gauged through the use of technologies that are most widely used for device-to-device connectivity. Initially, KHP smartphone app would connect to any Bluetooth-enabled device to collect device data.

Conclusions

Ultimately, traditional clinical information must be combined with genetic data and nontraditional phenotypes, and analyzed in a manner that yields actionable insights into disease diagnosis, prevention, or treatment. Real-time, quantitative human phenotyping and associated analytics will enable individuals, caregivers, and scientists to better quantify wellness and disease in a more continuous manner, and as individuals lead their daily lives. The bio-behavioral complex is the next great biomedical frontier, analogous to the Human Genome Project in its profound implications for medicine, and in the scale of the effort and resources required. With the knowledge

generated by CATCH and the KHP, we will finally close the gap between biological advances and healthcare practice.

Author Disclosure Statement

No competing financial interests exist.

References

1. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014;42:D1001–D1006.
2. Karlsson F, Tremaroli V, Nielsen J, Backhed F. Assessing the human gut microbiota in metabolic diseases. *Diabetes.* 2013;62:3341–3349.
3. Karlsson FH, Fak F, Nookaew I, et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat Commun.* 2012;3:1245.
4. Ley RE, Turnbaugh PJ, Klein S, Gordon JL. Microbial ecology: Human gut microbes associated with obesity. *Nature.* 2006;444:1022–1023.
5. Karlsson FH, Tremaroli V, Nookaew I, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature.* 2013;498:99–103.
6. Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature.* 2012;490:55–60.
7. Kostic AD, Gevers D, Siljander H, et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe.* 2015;17:260–273.
8. Cho I, Blaser MJ. The human microbiome: At the interface of health and disease. *Nat Rev Genet.* 2012;13:260–270.
9. Thanassoulis G, Peloso GM, Pencina MJ, et al. A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: The Framingham Heart Study. *Circ Cardiovasc Genet* 2012;5:113–121.

Cite this article as: Ausiello D, Lipnick S (2015) Real-time assessment of wellness and disease in daily life. *Big Data* 3:3, 203–208, DOI: 10.1089/big.2015.0016.

Abbreviations Used

CATCH = Center for Assessment Technology and Continuous Health
EMRs = electronic medical records
MGH = Massachusetts General Hospital