

QUANTITATIVE HUMAN PHENOTYPING: THE NEXT FRONTIER IN MEDICINE

DENNIS AUSIELLO, MD, and (*by invitation*) STANLEY SHAW, MD, PhD

THE OPPORTUNITIES OF THE GENETIC AND DIGITAL REVOLUTIONS

We are currently in the midst of an unprecedented profusion of powerful computing and communications platforms 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 applications (“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.

At the same time, advances in human genetics (especially genome-wide association studies) have identified unprecedented numbers of chromosomal loci that contribute to human traits and risk of human disease. These genetic data, combined with insights from basic, hypothesis-driven laboratory research, provide a much clearer outline of the genes that contribute to human 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 that may improve how diseases are diagnosed, prevented, and treated. However, the clinical translation of these discoveries into new approaches to diagnosis, prevention, and treatment has become the next critical challenge, and it faces several obstacles. First, the functional and medical significance of most susceptibility alleles is not well understood. Second, any disease trait is under the influence of not just many genetic loci, but also environmental and behavioral influences.

Correspondence and reprint requests: Dennis Ausiello, MD, Director, Center for Assessment Technology and Continuous Health, Physician-in-Chief, Emeritus, MGH, Bulfinch 127, Massachusetts General Hospital, Fruit Street, Boston, MA 02114, E-mail: dausiello@partners.org

Potential Conflicts of Interest: None disclosed.

The increasing availability of digital and genetic data has contributed to a third opportunity: the collection and analysis of extremely large data sets containing multiple types of data relevant to human health. These include the data contained in electronic medical records, genetic data (eg, 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 communications, 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.

CHALLENGES FACING HEALTHCARE AND BIOMEDICAL DISCOVERY

Our current system of delivering health care is episodic and reactive. That is, patients see their physicians largely at regularly scheduled intervals (typically 1 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. Whereas 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.

More broadly, the available methods used to diagnose and quantify disease have conspicuously lagged behind the recent exciting discoveries in the genetics and mechanisms of human diseases. Consider the cases of type 2 diabetes and lipid disorders, each of which is now associated with many dozens of chromosomal regions that influ-

ence 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.

Although 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 challenging at present to effectively use genetic information to stratify risk. In one 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). 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 NEED FOR NEW QUANTITATIVE HUMAN MEASUREMENTS THAT REFLECT HEALTH AND DISEASE

There is a fundamental need for new approaches to measure human health, so that we can better quantify wellness and disease in a more continuous manner, and as our patients lead their daily lives. A renewed focus on quantitative human measurements, or phenotypes, can benefit human health in numerous ways, from individual patient empowerment, to biomedical discovery, to new approaches to diagnosis, prevention, and treatment.

First, a new approach to human measurements can 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.

A related challenge is to make 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 (eg, 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 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 significantly increases the completeness of data captured and the populations to which this approach may be applied; active data entry risks selecting for participants who start with 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.

Novel phenotypes could aid biomedical discovery in several ways. 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 limited largely to existing phenotypes studied 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 non-small-cell lung cancer bearing mutations in the epidermal growth factor receptor or the anaplastic lymphoma kinase. However, for the majority of genetically complex, 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.

Ultimately, traditional clinical information must be combined with genetic data and non-traditional 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. This 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.

EXAMPLES OF NOVEL PHENOTYPING EFFORTS

At the Massachusetts General Hospital (MGH), we are engaging these challenges via the newly formed Center for Assessment Technology and Continuous Health (CATCH), which seeks to discover and apply new ways to quantitatively measure the human condition (eg, 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. In the paragraphs below, we describe examples of phenotypes that we believe are amenable to intensive research efforts. Because of the diversity of the universe of potential measurements, the development of integrative analytics that allow disparate data types (traditional and non-traditional) to be analyzed in concert will be critical.

Pathway-inspired Measurements

A number of pathways represent exciting new disease targets or novel genetic discoveries, yet are poorly addressed by existing human phenotyping. For instance, 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 that assess the activity of inflammatory cells (including molecular imaging of inflam-

matory cells, or microfluidic devices that can trap or analyze single cells). Shifting pathway-based measurements from the province of research studies to point-of-care devices for use in physician offices or the home would represent a significant leap in making these measurements more “continuous.”

Continuous and Behavioral Measurements

Certain types of medical data are already collected continuously that 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.

Another category of continuous measurements overlaps with behavioral measurements. Behaviors such as exercise, diet, and medication adherence 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 (eg, 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 used with high persistence, such as cellular phones or smart phones, may open up new avenues. Furthermore, because mobile phones are increasingly used for a variety of routine behaviors such as communication, travel location, and even specific health-related software (apps), the mobile phone represents an appealing platform for a variety of continuous and behavioral measurements.

Environmental Measurements

Finally, another group of poorly measured data includes environmental exposures, including diet, or potentially inhaled or ingested toxins. Exposures are typically accessed on rare occasions through survey instruments based on recall or blood or urine assays. Although

continuous measurement of environmental exposures may not be necessary (or feasible), enabling more facile, passive quantification of environmental exposures will create an important new data resource that can be integrated with genetic and clinical information.

Enabling Integrated Analyses

An example of the integrative potential of new phenotypes is 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 in some animal models, even behavior). One can easily envision 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. Integrating phenotypes at multiple scales of biology (from cells to the whole individual) will likely become the new norm in human disease studies.

A CHANGE IN CULTURE

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 are 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 with a 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.

And perhaps most importantly, we must collectively re-examine the culture of medical research and practice. Scientists, engineers, and care-

givers must partner with patients in new ways for a fuller understanding of individual health, as well as for presymptomatic, continuous assessment and care. The traditional barriers between clinical care and clinical research must be replaced by a new model in which patients are at the center as fully informed participants, and individual wellness is pursued hand-in-hand with a spirit of inquiry (Figure 1).

DISCUSSION

Longo, Boston: Thanks, Denny, for that interesting talk. I come back to this question. I know that we've talked about this. What is the definition of wellness? Is it a person who lives 100 years with no complex illness? Because we all have things that happen that are very transient and don't really change the natural history of most of our lives. So where do you pick the cutoff? Do you have to define wellness first, before you go and make these phenotypic judgments, or are you just talking about taking the appearance of a disease and trying to work back?

Ausiello, Boston: Well, that is indeed the key question. I would argue that, as practicing physicians and historically in medicine, we talk about wellness as a qualita-



FIG. 1. Changing the medical dialogue from disease to wellness. Daily life experiences will be captured and analyzed through a convergence of science and engineering. This analysis, which will be minimally invasive and minimally intrusive, will take place through public and private collaborations and feature a renewed partnership with individuals.

tive episode and yet have no real true measurement of it. One of the things that is coming out of this huge access to population-based information is that wellness, of course, is a bell-shaped curve. Some of it is in the eye of the beholder. But they are beginning to have metrics, and those metrics are not the standards that we have been traditionally associated with. For example a blood pressure of X, a blood sugar of Y. So I would posit that we need this kind of interrogation, first and foremost, to describe some concept of wellness and then, ultimately, the progression to illness.

Wenzel, Richmond: Thank you very much. My question is how do you envision integrating this system with prevention since obviously that would have the most impact? Do you see that as purely at the individual person level before they become a patient, or do you see any way to link that to unmet needs in effective public health policy and political will?

Ausiello, Boston: That's a wonderful question, and it's one that we think about a lot. I would again state that in order to make headway in this particular space, you need to begin with a group of elite performers. I don't mean that from a physiological point of view. I mean that from a participatory point of view. So we are building this activity into the interrogation of a primary care practice at Mass General called The Ambulatory Practice of the Future. It is designed to take care of faculty and employees of Mass General, tens of thousands of people. Needless to say, they are relatively loyal cohorts. And we need from that base to develop the kinds of dashboards and algorithms that you are alluding to that will transfer themselves into a community who have never heard of Mass General nor would have ever participated in these kinds of studies. I would say that population-based information — I would not have said this a year ago — but population-based information coming under the private sector is an important tool kit for us to partner with, not only because of the accuracy that seems to be developing from some of those assessments, but from the speed. I can tell you some anecdotes about that if you are interested.

King, San Francisco: My question has two parts to it. One is, you have been in education for a long time and I am wondering how you are thinking about educating us — the older community, but more importantly, the young people — to get this concept and to move it forward, and what changes do you think we have to make that happen? My other question is just more of a concern about a question. I am really worried that a lot of this will become commercialized and we will actually stop the progress that you are trying to make because people are going to want to try to make money out of it. So I wonder how you've thought about that.

Ausiello, Boston: Talmadge, I didn't put education up there because quite frankly it doesn't usually attract the kinds of resources. But it is absolutely essential that we build an educational program around this kind of virtual world that I am describing. Indeed, when you look at our young physician scientists today, they live in a virtual world in all aspects of their lives except medicine. I gave a talk to senate staff in Washington, and I said that we know more about the dating habits of America than we know about their blood. If we can begin to instill the same digital life cycle, if you like, that lives in the minds of these individuals, then as much as I love my institution at the Mass General Hospital, it is ludicrous to think that all of medicine and all of medical assessment needs to occur within the four walls of that institution. We are going to have to build robust virtual digital interrogation systems, as well as education systems, that run the gambit between the real world out there with our patients.

Czeisler, Boston: Thank you very much for this innovative glimpse into the potential future of medicine. I saw a fascinating presentation a couple of years ago by a group that was looking in assisted-living facilities at the circadian distribution of the timing of the occupants' activity, which rooms they were in and whether they were in their bedrooms

and so on. And that preceding deterioration and presentation of clinical symptomatology, there were abnormalities that could be easily detected in the circadian timing of the behavior. They were positing that this could be used, if this were being monitored on a large scale, to initiate much earlier interventions. Are you integrating that kind of thing into your proposed plan?

Ausiello, Boston: Yeah. I did not have time to go into the technology about the behavioral analytics and the presymptomatic assessment. You saw, however, those curves that we were able to predict. Now what is that about? We have been working with the media lab at MIT — and Sandy Pentland's group in particular — where they have developed smart phone technology, some of which has become popularized as a company called Ginger.io. But there is an academic form of it as well that is available to everybody where — both not only by the active participation but equally important by the passive analytics of the cell phone — you can begin to make predictions. So, for example, the average cell phone in the United States is never more than two and a half feet away from your body. It sits on your night stand, it gets up with you at 6 am every day, but 5 days in a row it gets up at 8:00. Instead of driving to the gym, you drive someplace else. If you're a student at MIT or Harvard, you call your mother three times in a week instead of once a month. These are the kinds of information assets that allow us to take deep dives of interrogation much sooner. From the studies with inflammatory bowel disease and with lupus, we are able to find one or two cytokines that go up a week before the plethora of immunological events that usually occur by the time we see them in a hospital. So you are absolutely right. The use of early behavioral analytics, both passive and active — and I don't have time to talk about the active ones that we are employing — I think, will be crucial for us to begin to parse this stratification that we are talking about.