



JOHN HARVARD'S JOURNAL

Isaac Kohane

Toward Precision Medicine

Building scale in biomedical informatics

ANTICIPATING “radical transformations” in medicine in coming decades, the dean of Harvard Medical School (HMS) has authorized a full-scale department of biomedical informatics, effective July 1. Jeffrey Flier’s move recognizes the growing importance of data in the healthcare professions, and, he said, builds on the school’s “outstanding record of achievement” in the field. Henderson professor of pediatrics and health sci-

ences and technology Isaac “Zak” Kohane will chair the new department. Since 2005, he has co-directed HMS’s Center for Biomedical Informatics (CBMI); five of its associates will become the department’s first core faculty members, and Kohane has committed to recruit 10 more colleagues during the next five to seven years.

What is biomedical informatics, why does it matter, and why now at HMS? The

emerging discipline “reflects the dramatic development of large data sets in genetics, genomics, studies of proteins, the nervous system—all aspects of biomedical science and ultimately patient care,” says Gilbert Omenn, M.D. ’65, director of the University of Michigan’s Center for Computational Medicine and Bioinformatics, who chaired the external committee that reviewed the proposal for the new department. But much of this information is “heterogeneous,” he explains: the data range from the molecular and genetic to the behavioral and sociological. “All of it,” Omenn says, “has to come together to paint a complete picture of the determinants of health and disease, as well as response to therapies and general care.” Biomedical informatics aims to create an information commons that will be useful to researchers, doctors, and even their patients. Kohane’s experi-

I N T H I S I S S U E

- | | | | |
|----|--------------------------------|----|------------------------------------|
| 19 | Harvard Portrait | 25 | Putting the “New” in House Renewal |
| 20 | Yesterday’s News | 27 | Brevia |
| 22 | Debating Sharia Law, Digitally | 29 | The Undergraduate |
| 24 | University News Briefs | 32 | Sports |

ence as director of informatics at Boston Children's Hospital, and at HMS's Countway Library (see "Gutenberg 2.0," May-June 2010, page 36), is directly pertinent.

Lessons from Netflix

"MEDICINE AS A WHOLE is a knowledge-processing business that increasingly is taking large amounts of data and then, in theory, bringing that information to the point of care so that doctor and patient have a maximally informed visit," says Kohane. He compares this idealized patient experience, in a sense, to Netflix or Amazon's connection to consumers: they already know "your entire prior purchase history...what other consumers with a similar history are going to buy next, and what to recommend to you." But in medicine, he points out, patients with chronic diseases must repeat some abbreviated version of their entire medical history "again and again to every provider."

To enable precision movie-picking or shopping, the titans of online commerce take advantage of "knowing a lot about a population and a lot about you," Kohane says. Most patients, on the other hand,

would be lucky if their providers knew about similar patients in their own practices—let alone all patients with similar histories—and about which drugs worked for different patient subgroups. This kind of precision medicine simply doesn't exist yet. (Kohane served on the National Academy of Sciences committee that delivered a 2011 report on the subject to President Obama; among his own projects is the creation of a central repository for information about neurodevelopmental diseases, with a special focus on autism spectrum disorder, in which both genes and the environment are thought to play important roles.)

Another goal of biomedical informatics is to improve diagnoses and treatments by removing some of the subjectivity of clinical interactions. For example, many doctors, when listening to a patient's heart through a stethoscope, "will disagree about what they are hearing, and what it means," says Kohane. "Whereas you can attach a computer to a microphone, and get consistent, reliable diagnoses of which valve is affected."

A related, pedagogical goal is enhancing physicians' search and numeracy

skills. "The best predictor of a doctor ordering a genetic test is knowing whether the patient asked for it," Kohane continues—usually because that patient has searched on Google. But researchers have found that even "well-trained physicians are both uncomfortable and incompetent in interpreting these tests," often because they lack numeracy skills. One of Kohane's former students, Arjun Manrai '08, asked doctors and residents at a Boston hospital a simple question: "If a test to detect a disease whose prevalence is one in a thousand has a false positive rate of 5 percent, what is the chance that a person found to have a positive result actually has the disease?" The test will yield 50 false positives in a population of 1,000, but only one patient will actually be ill—so a positive test result would mean that a patient has only about a 2 percent chance of having the disease. More than three-quarters of the respondents in the study got this wrong; the most common answer was 95 percent.

As Kohane puts it, "In the era where we're beginning to take away pieces of your body, like a breast, based on a ge-

HM Explore More

Harvardmagazine.com brings you continuous coverage of University and alumni news. Visit to find these stories and more:



Presidents Xi and Faust Confer in China

In Beijing, a conversation on climate change and academic collaboration

harvardmag.com/xi-15



Tim Mayotte Joins Harvard Tennis

The former Wimbledon semifinalist aims to improve the Crimson squad's level of play.

harvardmag.com/mayotte-15



Crunching the Numbers on Voting Rights in America

Why good data are essential to understanding the Voting Rights Act

harvardmag.com/voting-15

visit harvardmagazine.com



Commencement 2015

Planning for Commencement week exercises in May is already under way at the College and Harvard's graduate and professional schools. Read about the speakers who will be addressing the new graduates at Class Day and Commencement ceremonies at harvardmag.com/tags/commencement-2015

CLOCKWISE FROM TOP LEFT: KRIS SNIBB/HFAC; COURTESY OF TIM MAYOTTE; BEN DANNIER, JON CHASE/HFAC

netic test, we're going to have to better understand the meaning of probability." That points to the department's pedagogical mission: educating graduate students (who will become research scientists) and medical students in order to ensure that they know how to use the computational tools in increasingly wide use.

Grappling with Genomic Data

THIS NEW APPROACH is consistent with trends in medical education generally. "It's now accepted that in medical school you're only going to learn a tiny fraction of what you need to actually provide expert care for patients," Kohane points out—even in a narrow subdiscipline. Genomics, he adds, has compounded the problem by several orders of magnitude: "There's no way anyone, no matter how manic, is going to know what a million different [gene] variants mean for an individual." Doctors therefore need a computational "decision support infrastructure" that *interprets* a patient's medical history, family history, and genomic background to show what the risks are, as well as the preferred therapies.

Bringing individualized genomic information into the clinic will "accelerate the realization that we need just-in-time decision support," Kohane continues. But "process automation"—such as digitizing medical records to streamline hospital operations and doctors' offices—is nowhere near that capability, so tackling that problem is on the new department's agenda, too.

Such a system might do what Google has done for maps: layer atop a location's geographical coordinates all kinds of other useful information, such as current weather or crowd-sourced data on good places to walk or dine. Imagine if environmental exposures, genetic makeup, lifestyle habits, diet, and epigenetic information (about which genes are actually turned on or off) were "mapped" onto patient records. "Stacking that all together," he says, will provide "a better understanding of the patient as a whole" in order to predict, for example, if someone is at risk for diabetes.

New Models of Diagnosis and Care

TWO YEARS AGO, Kohane and colleagues demonstrated the power of integrating genetic data into diagnoses in a contest, the Clarity Challenge (see www.irdirc.org/?p=2892). Thirty teams of doctors around the world were given the histories

HARVARD PORTRAIT



Jelani Nelson

Jelani Nelson lights up when he talks about algorithms. The soft-spoken assistant professor of computer science is a rising star in a field made vital as data proliferate exponentially faster than the growth of computational power or storage. Algorithms, well-defined procedures for carrying out computational tasks, speed the way to answers. Nelson has a knack for speed: online, where he is known as "minilek"—a handle chosen in youth when he was growing up on St. Thomas, and derived from the name of an early ruler of Ethiopia, whence his mother hails—he has won with equal ease coding competitions and typing contests (topping out above 200 words per minute). Though he is a theorist now, solving real problems quickly "cements the concepts in your mind," he says. Borne of that conviction, every homework assignment in his undergraduate course Computer Science 124, "Data Structures and Algorithms," includes an algorithmic programming problem. His own student years were spent practically next door, at MIT, where he majored in mathematics and computer science, and remained to earn a Ph.D. in the latter field. He came to Harvard in 2013 after post-doctoral research at Berkeley and Princeton's Institute for Advanced Study. Nelson's specialty is "sketching," an approach to dealing with problems in which there are "too many data in the input." He figures out how to create compressed, often exponentially smaller, versions of datasets that nevertheless retain useful, accurate information. His proofs defining the limits of such approaches have illuminated fundamental questions, some of them unanswered for decades. Though he is humble and quiet, his colleagues are less reserved: they call him "simply brilliant." —JONATHAN SHAW

Yesterday's News

From the pages of the *Harvard Alumni Bulletin* and *Harvard Magazine*

1925 The senior and freshman classes assemble on the Widener steps to have their respective pictures taken, and the freshmen, according to custom, contribute to the seniors' pre-Commencement celebration: the total is \$244.24 and two cats, one alive and one dead.

* * *

Thirteen Harvard men gather on May 1 to organize the Harvard Club of Shanghai; Way Sung New, M.D. '14, is elected president.

1940 After Germany invades the Netherlands, President Conant argues on national radio that "the changed military situation in Europe threatens our way of life," and student support for aid to the Allies increases.

1950 Harvard enjoys its "most amazing financial year in history," raising nearly \$26 million in gifts, bequests, and grants.

1955 The Corporation approves a new doctoral gown for Harvard degree holders, "crimson silk and worsted stuff"

faced down the front with black velvet and with three black velvet bars on each wide bell-shaped sleeve.

1980 Class Day speaker Walter Cronkite warns graduating seniors that unless they come to grips with the "mega-problems" of overpopulation, pollution, natural-resource depletion, and nuclear proliferation—"our modern Four Horsemen of the Apocalypse"—civilization as we know it cannot survive.

1990 In mid May, President Derek Bok makes public the Harvard Corporation's eight-month-old decision to remove from the University's portfolio all stock in firms that manufacture tobacco products.

1995 Class Day speaker Hank Aaron shares a story about a young man who went running up to his father, saying, "Look, Dad, I got it! I got my A.B. from Harvard." To which the father replied, "Son, that's fine. We are all real proud. Now it's time for you to go to work and learn the rest of the alphabet."

and genome sequences of three families, each with a sick child whose disease was believed to be genetic. Seven teams converged on very useful diagnoses for two of the patients (including a case that had gone 11 years without one), a result that demonstrated the potential utility of genomic data in such cases. But there was an even more significant lesson, Kohane believes: one of the patients, he reports, "had *already been evaluated* in two of the hospitals that were home to winning teams," but had remained undiagnosed. During the challenge, on the other hand, the teams successfully identified the problem. That means, he says, "that there is a *better* process, that does not look like the *current* process, of medical care that is multidisciplinary, and involves the use of computational experts, as well as genetic experts, as well as clinicians, working as a team to create qualitative—or quantum—differences in care."

That experience persuades him that the new department will generate "new models of diagnosis." The HMS biomedical informatics group has already been tapped as the coordinating center for a new national network on undiagnosed diseases. This program, based on a pilot developed at the National Institutes of Health that resulted in successful diagnoses for rare diseases 30 to 50 percent of the time, involves genome sequencing and then a referral for the patient to the leading specialist in the disease.

Ultimately, Kohane says, such efforts will require both a new infrastructure that can encompass all kinds of, and *lots* of, data, and a new kind of caregiver: individuals who "want to make a difference in biology and medicine and yet are wizards in quantitative reasoning and computational methods." In his estimation, such "quants" could probably have greater impact than any single doctor by identifying early signs of disease, finding new treatments, and warning about drugs that do not work.

A New Breed of Doctor

CHIRAG PATEL, a CBMI research associate in biomedical informatics who will become a faculty member in the new department

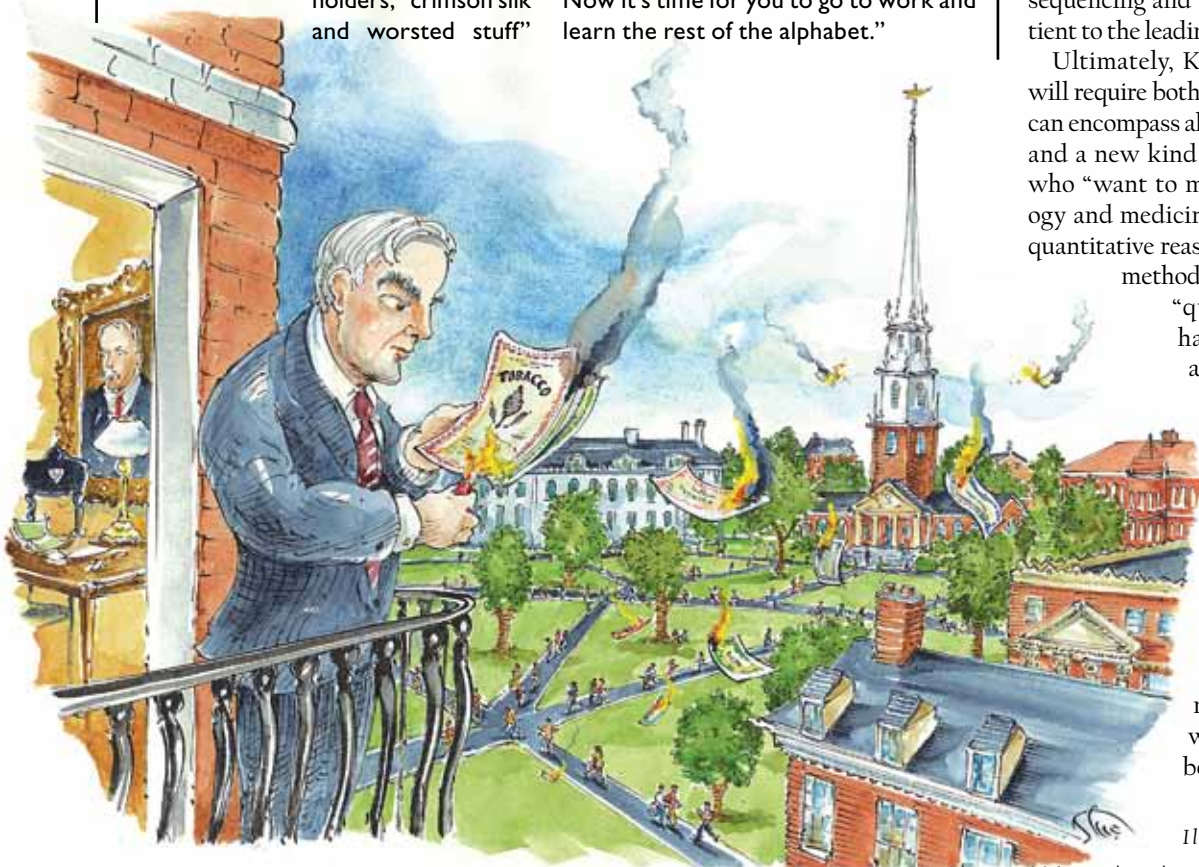


Illustration by Mark Steele

in July, is one role model for this new generation of student. After studying molecular biology and computing at the University of California, Berkeley, he began working as a software engineer at a biotechnology company focused on genome sequencing. Later, mentored by biomedical informaticist Atul Butte at Stanford, Patel “fell in love” with biomedical informatics because it combined genomics with computer science, statistics, and mathematics—“all my core interests.” After earning a Ph.D., he worked for a year with Stanford professor of medicine John Ioannidis (who is also an adjunct professor at the Harvard T. H. Chan School of Public Health), developing analytical methods to mine large, epidemiological data sets.

One of Patel’s current projects is to develop software for environment-wide association studies that will allow researchers to study the relationship between genomes and exposomes (the totality of a person’s environmental exposures to such things as drugs, diet, diseases, and pollutants). For example, type 1 diabetes, an autoimmune disease, arises spontaneously, typically in children and young adults. Researchers know that some people have a pre-existing genetic susceptibility, but also that the disease is *triggered* by an environmental exposure. Figuring out which exposures cause the autoimmune response has proven challenging: an individual genome represents a huge amount of data, but at least it is a discrete entity. A person’s exposome, on the other hand, encompasses data ranging from electronic medical records, to membership in epidemiological cohorts, to personal exposure monitoring—a *gigantic* “big data” problem (see “Why Big Data is a Big Deal,” March-April 2014, page 30).

Fortunately, Patel “likes to look at everything at once. I try to hammer out connections...and correlate everything with everything else.” Then he tries to “fish out the signals from the noise.” Correlations will always emerge from enormous sets of data; the challenge is figuring out which exposures have a basis in biology—and merit further exploration. For example, Patel and other researchers have found a link between diabetes and such persistent pollutants as the pesticide DDT and polychlorinated biphenyls (PCBs, commonly used as coolants in electrical apparatus and other applications). “But we still can’t pin down the biology behind these correlations,” he explains. “Are they biased



Chirag Patel

by other factors such as age, for example, which is a huge risk factor for diabetes and cardiovascular disease?”

Training students to ask the relevant questions, Patel believes, is one of the biggest challenges in biomedical informatics. Students need to be just as adept in biology and patient care as in advanced computing and data analysis. “The challenge is finding folks who can bridge those worlds.”

Forming a Faculty

RECRUITING additional faculty members who can do that is Kohane’s responsibility. He seeks expertise in three areas. The first involves creating a *patient-centered information commons* that will bring together patient data of diverse kinds so that it can be incorporated into population-level research, and make the findings useful in individual care. Such a system will demand rapid computation across millions of individuals and hundreds of thousands of data types, as well as privacy protections. Scientists working on this effort, including Patel, come from the medical and public-health schools and the Harvard/MIT Health Sciences and Technology program (the training ground for biomedical researchers). The aim is to enable precision medicine that could, say, combine data from a patient’s psychiatric history, genetic information, and records of environmental exposures to derive clinically relevant information.

His second focus, noted earlier, is identifying faculty members who can develop new tools and techniques for using data to *generate automated diagnoses* more accurate than those the current healthcare system provides.

The third priority involves *reimagining the clinical encounter*: “How do we provide the providers with all that just-in-time data about the patient? How do we provide it in a way that is useful in making decisions about the patient? How do we allow them to measure things quantitatively—through noninvasive imaging such as ultrasound—and integrate that into their assessment of the patient?” The kind of person who could successfully change the healthcare encounter will probably combine skills in systems engineering, human-productivity and effectiveness engineering, and a variety of real-time information technologies.

These challenges are immense, but the effort to put such knowledge back in the hands of healthcare providers is overdue. As Kohane puts it, “For a variety of reasons—some out of reasonable caution, but some out of institutional inertia—medicine has been slower than other disciplines to take advantage of the new insights and the new productivity that you can get through data science and process automation.” His department aims to make a difference throughout biomedical research—and practice.

—JONATHAN SHAW