FROM EVOLUTION TO REVOLUTION:

Building the 21st Century Genomic Infrastructure

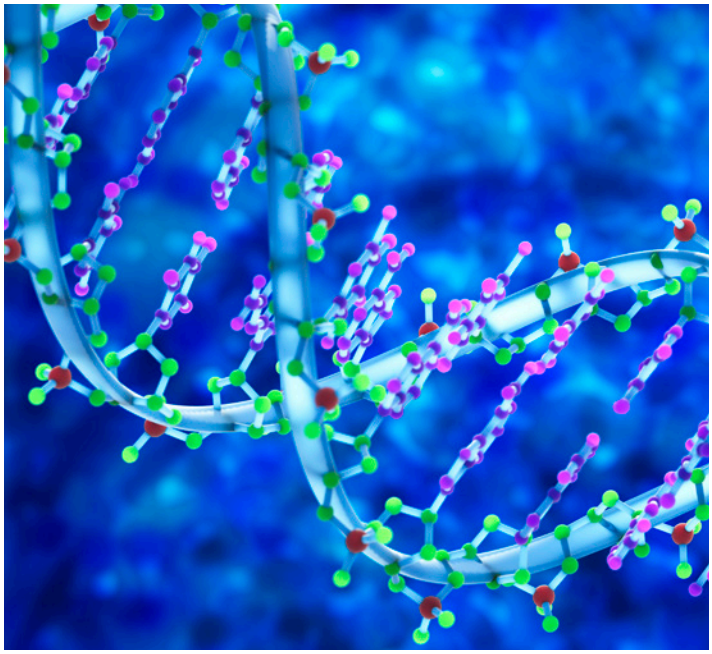
I. EXECUTIVE SUMMARY

Much has happened in the 15 years since the sequencing of the human genome was unveiled in the East Room of the White House. President Bill Clinton likened that first sequence to Lewis and Clark's map—the trail that, legend has it, launched the nation on the collective enterprise of westward expansion. The pioneers of the Human Genome Project, likewise, realized that sequencing had, for the first time, sketched out the frontier of human biology, pointing the way to an era of precision medicine—humankind's new manifest destiny. What they could not have known was how difficult it would be to translate their early insights into clinical know-how, or the sheer power of the tools needed to finish the job.

President Barack Obama launched the Precision Medicine Initiative during his 2015 State of the Union speech to further accelerate our understanding of individual variability and its effect on disease onset, progression, prevention, and treatment. To assist this effort, the president included $215 million in his latest budget to fund initiatives at the National Institutes of Health (NIH), the National Cancer Institute (NCI), the Food and Drug Administration (FDA), and the Office of the National Coordinator for Health Information Technology (ONC), including:

- $130 million to NIH to develop a national research cohort of a million or more volunteers to provide a large genetic research pool and pave a way for completing research via engaged participants and open, responsible data sharing.
- $70 million to NCI to scale up efforts to identify genomic drivers in cancer and apply that knowledge to the development of more effective treatments.
- $10 million to FDA to develop high quality, curated databases needed to revamp the regulatory structure to support precision medicine while protecting public health.
- $5 million to ONC to support development of interoperability standards and requirements that address privacy and enable secure data exchange across systems.

Given the promise of precision medicine to deliver cures to specific subsets of disease populations, including rare diseases, the House and Senate have made this initiative a key priority. To be maximally effective, however, this initiative will require the public and private sectors to work in tandem to realize the next generation of medicine and health care and overcome the institutional challenges that increasingly hinder progress toward precision medicine. This will require policymakers to modernize the regulatory system to better enable big data analytics to be used in genomic research. To that end, we recommend the following:



1. Improve interoperability and data sharing. Stronger federal requirements are needed to ensure that genomic and other health data can be retrieved and compared across health record systems. Bottom-up, patient-driven reforms, such as giving patients (and their providers) a right to access and share interoperable health data, would incentivize standard setting and save lives.
2. Engage patients. The public and private sectors share an interest in raising the tone of discourse on the role that genomics and other big-data applications might play in revolutionizing our expensive and underperforming health system. As the true customers of health care, consumers—especially patients desperately waiting for breakthrough cures—must be brought into this dialogue.
3. Re-think privacy law. The strict privacy requirements of the Health Information Portability and Accountability Act and complementary federal and state laws, including the Common Rule, present formidable obstacles to realizing the potential of genomic medicine. It is time to reassess the costs and benefits of these policies in light of recent scientific and technical advances, and to consider less burdensome models for protecting privacy. This white paper is the product of a December 11, 2014, symposium in Washington, D.C., co-hosted by Health IT Now and the Center for Data Innovation, to examine the role of big data analytics in genomic research. It is intended to assist policymakers as they dig through the policy and practical issues associated with precision medicine, and provide actionable policy options as Congress and the Administration seek to foster precision medicine treatments aided by innovative technologies and data.

II. INTRODUCTION

We now know the genome to be remarkably more complex than was understood in 2000 and that its reverse-engineering requires the intensive application of data management, computational, and observational tools that only recently have come into use—and which are still rapidly improving. By some estimates, U.S. health providers have spent more than $100 billion upgrading (or simply adopting) information technology since 2009.[1] Public and private institutes, universities, pharmaceutical manufacturers, and biotech startups have spent many tens of billions more plumbing the depths of the genome. Yet these expenditures could pale in comparison to the needed investments of the next decade.

As the technical limits to genomic research recede, institutional barriers are looming larger. Unlike the open-source, international collaboration that confirmed the existence of the Higgs Field, biomedical research is cramped by compartmentalization that depletes research dollars and slows the pace of discovery. Many billions are wasted on outdated clinical trial protocols that fail to make use of predictive toxicology and other tools made possible by genomic medicine. In addition, regulatory obstacles diminish the private-sector business case for building research infrastructure and partnering with the public sector. Clearing this regulatory underbrush can hasten the development of cures for the half or more of diseases that are untreatable today—including 95 percent of rare diseases.[2] Importantly, regulatory reforms can help shore up America's eroding preeminence in a field that promises to transform the 21st century.

III. DISCUSSION

Hardly a day goes by without breaking news on the genomics front. Sometimes it is sensational, as in the U.K.'s approval of an artificial insemination technique that uses donated mitochondrial DNA to prevent a devastating inherited condition.[3] More often, it is the background thrum of reorganization sweeping the pharmaceutical industry, such as AstraZeneca's licensing of the CRISPR genome-editing technology to develop cancer cures.[4] And sometimes it is just plain confounding. In this category is the recent finding that genes from synergistic bacteria in our guts—the microbiome—get imprinted onto our genes and passed along to our children.[5] This complements a 2009 finding that up to eight percent of the human genome is derived from viruses.[6]

A. Building the Knowledge Base

The genome is proving to be a multidimensional puzzle. Of its roughly three billion base pairs, only 1.5 percent are coding elements capable of producing proteins, fitting the classic definition of "genes." The prevailing view in 2000 was that the remainder was "junk" that had outworn its usefulness. Researchers consequently assumed that most genomic medicine would involve manipulating the 23,000 coding genes—the exome. That was too simple.[7]

Subsequent studies have shown that 95 percent of genetic material consists of noncoding RNA, complex molecules that chemically resemble DNA but typically have one strand instead of two. RNA comes in many varieties, however, and unlike DNA, whose double helix hardwires it against destruction, many RNAs blink into and out of existence, making them difficult to study. Most are messengers carrying instructions and building materials to genes and proteins. But some turn off individual genes or blocks of genes. Still others mediate proteins, called transcription factors, which regulate replication.[8]

Most traits and many complex diseases originate in regulatory regions of the genome that do not code for proteins. Biologists believe that variations in RNA may solve the riddle of "missing heritability" (for example, the fact that common diseases such as type-2 diabetes and Alzheimer's are not explained by combinations of gene variants).[9]

These insights are slowly working their way into the clinic. Sovaldi, a cure for hepatitis C, specifically targets the enzyme NS5B polymerase, which regulates the hepatitis C virus RNA replication.[10] Another therapy enhances cellular receptiveness to the cancer drug Taxol, by commandeering a common virus that does not cause illness (humans carry about 100,000 such viruses) to deliver beneficial instructions to tumor cells. The active ingredient, in this case, consists of tiny RNA fragments about 21 letters long.[11] A promising line of research on HIV vaccines uses a similar technique to manufacture molecular decoys that distract the HIV-1 virus (one of seven variants) from attacking immune cells.[12]

Other practical applications come from the field of pharmacogenomics, the study of genetics in drug response. Each individual's ability to absorb, distribute, metabolize, and eliminate drugs is determined by genes. Genetic variation may explain why Benadryl, an antihistamine, makes some people sleepy and others jumpy.[13] With enough information on which genetic combinations lead to what responses, an individual's responses to medications can be predicted, down to the

appropriate dosage level—all without the trial and error that plagues today's medicine. (Many drugs have failed clinical trials due to overlapping efficacious and toxic doses. These now can be managed with personalized treatment algorithms.)

Parallel advances in laboratory tools are also helping to drive the search for therapies. One of them is optical super-resolution fluorescence microscopy, first introduced in 2008. This technology permits 3D observation of the tiniest components of living cells, such as RNA, down to one-quarter of the wavelength of light—a level of detail once thought impossible.[14] Another is gene sequencing, whose cost has fallen from roughly $95 million in 2001[15] to $1,000 in 2015,[16] and could soon fall to $300 or lower.[17]

Underlying all of these advances has been the stunning explosion of computing power.

B. Building the Infrastructure

All of this might qualify as amazing progress, but for one fact: It has produced only a trickle of treatments. Roughly 150 gene therapies are in use today, but most of them—such as Xalkori, approved in 2011 to treat the four percent of non-small-cell lung cancers driven by rearrangements in the ALK fusion protein gene—are narrowly focused and insufficient to produce cures (in part, because cancer genomes continue mutating). Many thousands more are needed. Meanwhile, progress on common diseases, such as type-2 diabetes, is hindered by heterogeneity, a phenomenon where the same disease can result from many different mutations in many different genes (the "many roads to Rome" problem). Exploring this unknown territory will require the application of big data analytics on a scale almost unimaginable in 2000.

The genomic code is itself a repository of big data. With three billion nucleotides (chemical "letters"), a whole genome sequence, or WGS, takes up 150 gigabytes of computer storage space (not including a backup)—the equivalent of 100 feature-length movies. So large are the memory and bandwidth needs that bulk sequencing vendors, such as the Beijing Genomics Institute (BGI), transmit their finished products on disk drives, via snail mail. The cost of storing a single WGS in the special media needed for ultra-rapid processing currently runs about $540 per year.[18] Meanwhile, the complex picture that is emerging suggests that genomes from a great many individuals—perhaps millions—must be mapped, catalogued, and compared before the full extent of human variation comes into focus.

Yet comparing genomes by employing high-level computing is only the first step. Interpreting the physiological role of variations in the exome, transcriptome, and proteome (the latter reflecting protein composition, structure, and activity patterns) requires correlating each patient's genomic variations with their conditions, symptoms, and drug responses—a body of evidence called the phenome. And even that may not be enough: The epigenome, which includes environmental factors, such as cigarette smoke, and the bacteria, fungi, and viruses that cohabit our bodies, is also known to trigger genetic mutations.

Patient health records offer a rich trove of phenomic data. Still more will come from a growing array of mobile health devices and monitors that make up the medical "Internet of things"—both do-it-yourself (non-medical) and FDA-approved.[19] Yet only recently have health providers begun recording patient information electronically. Given the growing digital saturation of everyday life, it may surprise some to learn that much of the health-care industry is digitizing from scratch. As recently as 2008, only 9.4 percent of hospitals had a basic electronic health record (EHR) system. This is rapidly changing. In 2014, the total had jumped to 75.5 percent, in part due to Medicare and Medicaid reimbursement incentives.[20] But even now, the most valuable information in the EHR consists of unstructured analog data, in the form of typed or scanned handwritten notes. Currently, the only way to use this data is to read through it—a time-consuming process. Converting this data into digitized, interoperable formats is computationally intensive—and to be reliable at scale, requires further advances in language recognition software.

Until now, the high cost of sequencing has kept a lid on use. But costs have plummeted so rapidly—declining by a third over the last year alone—that demand is now beginning to respond. A widely cited 2013 report projected that five years hence, the annual increment to genomic data would be enough to fill a stack of DVDs stretching all the way to the International Space Station. This may be conservative. Sequencing is still relatively inefficient (genes must be sequenced in millions of overlapping "reads" of a few hundred letters and then reassembled 30-40 times using statistical algorithms). There is no reason to believe that costs will not fall further, generating more demand. Going forward, the limiting factor could be the technical infrastructure itself.[21]

Thanks to Moore's Law (named after Gordon Moore, who predicted in 1965 that the number of transistors on a circuit board would double every 18 months), aided by the powerful multiplier effect of better algorithms, data storage and computing capacity has exploded in recent years, fueling an era of digital plenty. Today distributed, cloud-based computing now makes it possible to analyze in seconds or minutes what would have taken weeks or months in 2000.[22] But even with these continued advancements, the rapidly falling cost of gene sequencing suggests that there may be a computing and storage crunch on the horizon.[23] Add these infrastructure needs to the cost of converting years of paper health records into searchable electronic formats, and the bill for health IT over the next decade alone could easily run into the hundreds of billions.

C. Building the Business Case

Because genomic medicine will keep people healthier, investing in the necessary technology should make economic sense. Unfortunately, the business case for upgrading the IT infrastructure is not straightforward. If the results of genetic tests are not embraced by those who prescribe, interpret, pay for, and regulate them, the demand for genomic research will be limited. And while health spending in the U.S. will total roughly $42 trillion over the next decade, much of this money will not be spent on technological upgrades, but rather is likely to be squandered on unnecessary care, neglect, and mistakes, according to the Institute of Medicine.[24]

One problem is that many of the predicted paybacks would come outside the clinic. Investments made to improve adherence to medications keep patients healthy and out of the hospital, which translates into savings on the medical or clinical side of the ledger.[25]

Currently, the information from a healthy patient's genome offers only modest immediate clinical benefit, though the research value can be considerable. FDA has approved genetic tests only for specific conditions. Standards developed by the American Medical Association and adopted by FDA, meanwhile, bar the use of test results not ordered by a physician in clinical decision-support applications. Nor does it help that only a tiny percentage of physicians are trained in genetic interpretation. Lags in training may help to explain the medical profession's notoriously slow take-up rate: It can take 17 years for a new idea to work its way into the clinic.[26]

For their part, payers are concerned about the costs associated with diagnostic testing, both for the test itself, and potential demand generated by the test's results. Insurers worry that patients might demand costly follow-ups to see whether their biomarkers (such as variations associated with Alzheimer's) are causing imagined, or simply untreatable, health problems. Wasteful medicine already exacts a hidden toll on consumers, accounting for one-quarter or more of insurance premiums—$4,200 this year for a family of four.[27] The possibility that widespread sequencing might increase rather than decrease costs needs to be taken seriously.

Finally, the ability of drug companies to launch the hoped-for era of hyper-innovation depends on their financial health. Pharmaceutical research and development spending ballooned nearly five-fold during 2000-2007, but in recent years has fallen relative to inflation. The late 2000s also saw a 40 percent decline in the annual number of new drug approvals. One widely cited analysis combines these trends to project that the cost of bringing a drug to market (including failures) now tops $2.5 billion, up from about $800 million in 2001.[28] Some argue these trends and falling revenue from the "patent cliff"—the result of multiple blockbuster drugs coming off patent—will dampen the steam powering the locomotive of genomic invention.

This view may be too pessimistic. The drug development pipeline is a long one, with this year's approvals reflecting decisions made a decade or more ago. By this reckoning, the falling approval rates in the late-2000s were attributable to hypotheses of the pre-genomic era. Likewise, increased R&D in the early 2000s would have driven up average costs in the short term.[29] Lately, though, the results have been encouraging: 2014 saw 41 new drug approvals, 20 percent of which were related to precision medicine.[30]

Contributing to this improvement have been changes in the way drug applications are made and considered. Increasingly, drugs and diagnostic tools are being developed hand in hand. For example, Pfizer's cancer drug Xalkori was approved in tandem with a test kit developed by AstraZeneca that identifies which patients will respond. FDA, in turn, has used its recently expanded approval authority to move promising medications for serious illnesses directly into investigational therapy based on proven effects on biological activity (referred to in the medical community as "surrogate endpoints"), rather than evidence of tangible clinical outcomes which require expensive, lengthy large-scale randomized clinical trials. The pharmaceutical industry's ideal is a "quick win/fail fast" model, in which the reduced dangers of side effects and inappropriate prescribing lead to early point-of-care studies and much improved success rates. Development costs might be reduced 30 percent this way, says one industry report.[31] A 2012 report by the President's Advisory Council on Science and Technology (PCAST) calls for more regulatory innovation to this end, including greater reliance by FDA on predictive toxicology.[32]

The problem until now has been too few good drugs to approve. If translational research succeeds in identifying many thousands of promising molecular targets, this calculus would be turned on its head. Both FDA and the pharmaceutical industry will need to scale up their capacity for throughput; otherwise, much of the industry's growth will come from overseas.

Perhaps the strongest business case for building out the genomic research infrastructure belongs to the public. Federal and state programs cover half of all health bills (and thus would benefit disproportionately from cost-saving technologies), and many of the indirect benefits of genomic medicine, from fewer incurable diseases to stronger economic growth, are classic public goods. An apt analogy is to the highway system: The private sector doubtless would have built roads in the absence of public spending, but it likely would have taken longer and been more difficult. It is unfair, and unwise, to expect the private sector to shoulder the burden of a genomics research infrastructure alone.

Nor is the rest of the world standing still. China's public sector now spends twice much on translational research as we do. BGI has used this funding stream to acquire one-third of the world's sequencing capacity. Leading drug companies from the United States and Europe are opening research facilities near BGI-Shenzhen, to capitalize on its knowledge.[33] Meanwhile, weak investment and low research productivity in the United States are prompting researchers to look abroad for opportunities, creating a reverse brain drain. The risk that we will lose our lead in this defining technology of the 21st century to international competitors with fewer entrenched interests and lower regulatory costs is real and growing. The disruptive cures will still come, only they will not be made in America.[34]

IV. RECOMMENDATIONS

For the reasons outlined here, we believe that public sector leadership of a project to sequence and analyze the genomes of a million data donors-the Precision Medicine Initiative- would do much to spur innovation and create a business case for the build-out of America's genomic research infrastructure. A million-genome project would draw more private investment, and more competitors, into the development of precision medicines. It would generate new demand for the storage, standardization, and processing of data, along with associated software applications. These would be welcome developments. But spending more on research is only a partial answer. Also needed are reforms that improve the productivity of research.

Like the genome itself, the health system is swathed in regulation. Federal and state rules govern everything from the scope-of-practice of 1,100 different licensed health professions to every facet of the transfer, analysis, and communication of health data. These strictures lock in place paper-era practices and shield medicine from the kind of disruptive, IT-driven innovation that is transforming other sectors of our economy. One effect has been to channel invention toward high-cost, marginal improvements to old ways of doing things (think robotically assisted surgery or proton-beam prostate therapy).

Moving from an evolutionary to a revolutionary model of change requires regulatory innovation as well. To this end, we offer three avenues of reform:

Improve Interoperability

Compartmentalization occurs when EHR systems require users to purchase proprietary software. When physicians in one hospital lack access to records from another, the usual result is waste and duplication. But in some small percentage of cases, it will cost patients their lives. Interoperability has been a top-down process until now, with the federal government issuing rules and EHR vendors and their customers finding ways around them. These rules need to be strengthened. For example, the 21st Century Cures bill now pending before Congress would require adoption of industry-developed interoperability standards and make interoperability a condition of EHR certification beginning in 2019. But Congress should also explore consumer-driven, bottom-up solutions.

Many health-care providers today use EHRs courtesy of the roughly $30 billion in incentives authorized by the Health Information Technology for Economic and Clinical Health Act of 2010 (HITECH). In addition to grants, HITECH put in place "meaningful use" rules designed to specify minimum levels of technology functionality. These rules have been delayed and watered down, in part, to accommodate the self-proclaimed technical illiteracy of medical practitioners—most of whom had not adopted even basic EHRs at the time HITECH was enacted.[35] Indeed, there is often no business case for hospitals and physicians to modernize their practices, since they profit from waste and inefficiency in the health-care system.[36]

The Precision Medicine Initiative provides policymakers with a welcome opportunity to reboot. Its implementation will require universally accessible EHRs coupled with scalable technology, at least among the million-genome cohort. At a minimum, strong interoperability requirements are essential. These should include standardized vocabulary and data, and standards-driven protocols for sending, receiving, and querying records. Unique patient identifiers (in lieu of Social Security numbers) would provide a simple, uniform solution to identify patients across the health care system, linking them with their health data, and easing the administrative burden of managing a host of non-interoperable, proprietary identifiers.[37] Congress regularly precludes adoption of unique identifiers, and thus acts as the most significant barrier in this area. Congress should allow identifiers to proceed, recognizing the trade-off between use of data and safety and cost. By adopting such innovations, the Precision Medicine Initiative could drive change in the broader medical industry.

Congress should also look for ways to strengthen "bottom-up" market signals. One sensible approach would be to give patients co-ownership of their EHRs, including their sequencing results. Making patient health data a property right would leave no doubt about patients' rights to donate their data whenever, and to whomever, they see fit—and to have their genome analyzed periodically in light of the latest research. Some have dubbed this concept "patient-directed APIs" (short for application programming interface), where patients direct access and use to their information based on interfaces within their control.

Furthermore, providers should be required to allow patients real-time access to the data in their EHRs using industry standards. Where such interoperability does not exist, physicians, hospitals, and testing centers would have to provide it at their expense. This approach would update an existing right, under the Health Insurance Portability and Accountability Act of 1996 (HIPAA), for patients to be given access to their health records within 60 days of requesting it—and strengthen the market for vendors with compatible products.[38]

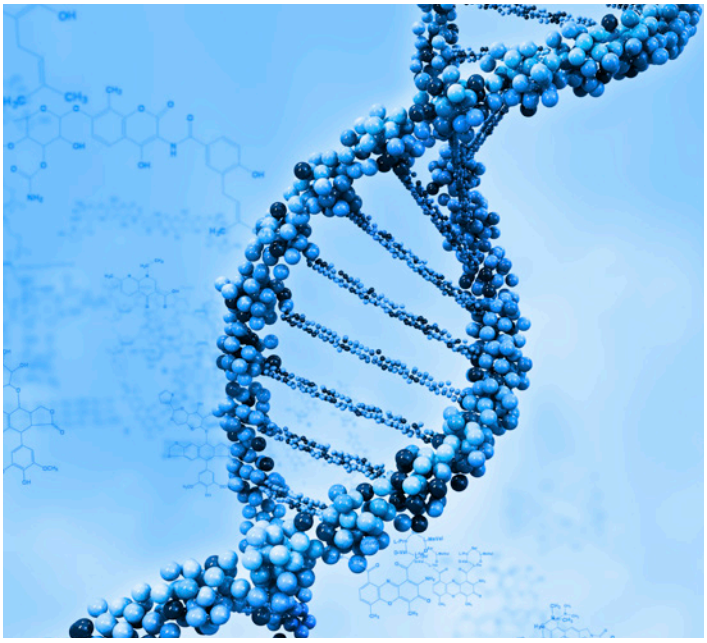
Outside of clinical care, interoperability problems vex the research community as well. One example is the lack of standardization in consent forms. Standardization would help foster the scalability of genomic research data—what is often referred to as "big data." Currently, the many federal and institutional regulations that govern what goes into a consent process constrain research data use. Short of requiring a universal consent form, Congress should require that all consent forms be created in a machine-readable format to allow researchers to more easily integrate new data into their studies while respecting patient preferences.

Engage Patients

For all of the trillions spent on health care, remarkably little goes toward communicating with patients. This is, in part, because, in the medical business, the insurer is the paying customer. It is also the legacy of an era when providers were thought to have all of the answers and meddling by insurers or government was viewed with suspicion. Data is changing this. As medicine moves into the 21st century, so, too, must the dialogue.

Doctors are no longer the experts they once were. IBM calculates, for example, that physicians would have to read 160 hours a week in order to keep up with medical developments.[39] Anticipated advances in bioinformatics—a field that most physicians only vaguely understand—could leave practitioners gasping to keep pace. If Americans are to have cutting-edge medicine, artificial intelligence, driven by big data, will need to play a greater role. In the future, health care is likely to be two parts data, one part doctor.

For their part, consumers have grown accustomed to researching health issues online. Rare disease networks have sprung up on Facebook and other social media. Americans are creating their own data points through mobile trackers, such as Fitbit. Until recently, 23andMe, a gene-sequencing vendor, did a brisk business providing biomarkers to those keen to learn more about the health implications of their genetic makeup—earning *Time Magazine's* "Best Invention of 2008" rating. In 2013, FDA (patronizingly) banned such services, to protect consumers from possibly misleading results (but recently has relented in narrow cases).[40] The widespread embrace of 23andMe is emblematic of a culture where data is more pervasive, and less threatening, than it once was.



For sequencing to move into the mainstream, there should be rewards for those willing to share their essence with medical science. Consumers who pay for their own sequencing often do so to obtain health information. Data donors should have access to basic information on the health implications of their genes along with the right to share their sequences with third parties for periodic review in light of new findings. Another model is provided by the Red Cross, which provides small financial rewards to blood donors. Employers might provide data donors with a discount on their health premiums, much as they do for participation in wellness programs. Researchers and others might provide direct transportation or subsidies to help get participants to and from trial settings or to fund the costs of missed days at work or for child care.

The public and private sectors share an interest in educating consumers about the role that genomics and other big-data applications might play in transforming our expensive and underperforming health system, as well as the safeguards already in place to address potential risks. Many consumers are unaware, for example, that since 2010, insurers have been barred from using health data to deny coverage or raise rates; or that, since 2008, employers have been prohibited from using genomic information in hiring. For their part, researchers whose livelihoods depend on their access to data are no more willing to self-destruct than drivers are to swerve into opposing lanes. Clear lines must be drawn against careless or objectionable practices, but the rules of the road that make other social interactions largely self-governing apply to medical research, too.

As we have seen in so many other domains (including the app-laden devices in our pockets) data is far more likely to liberate consumers than oppress them.

Re-think Privacy

The prospective benefits from genomic medicine merit a critical review of the opportunity costs of privacy rules. Under HIPAA and myriad supporting state and federal rules and statutes—including the Common Rule, which applies privacy standards to research subjects—health records are treated akin to military secrets.[41] Large-scale genomic studies can be conducted only in tightly controlled, "need to know" environments using (in theory) encrypted, statistically de-identified data sets. Violators are subject to severe civil and criminal penalties. Each use of data requires separate clearance—and in some cases, separate consent by each individual whose health data is being reexamined.[42] While these rules are intended to protect patient privacy, the same rules prevent drug companies from identifying or contacting those patients most likely to benefit from clinical trials. The effect has been to discourage collaborative, multidisciplinary research, divert scarce research dollars toward lawyers and cybersecurity, and drive up the cost of drug development. Arguably, lives are being lost in the bargain.

The president's million-genome project will need to address these issues head on. Discovering the phenomic manifestations of genetic variation will require the correlation of all data donors' genomes with their health records and other personal information. Unless provisions are made for the active involvement of hundreds of institutions and many thousands of researchers, discoveries will follow too slowly to benefit U.S. competitiveness—or to help the millions who suffer from untreatable illnesses today.

The genome is a molecular fingerprint that identifies not only individuals but the heritable traits of their kin. Short of hermetic compartmentalization, perfect security is impossible. To prove this, biostatisticians at Massachusetts Institute of Technology used genetic clues to identify roughly 5 percent of participants in the 1,000 Genomes Project.[43] In other words, donors to the Precision Medicine Initiative should understand that it is possible that some portion of their data may be re-identified, and in the battle against many incurable diseases, this is a trade-off many may be willing to make. As a result, donor consent forms will need to override state privacy laws that mandate tighter restrictions.

Working privacy alternatives to HIPAA are found throughout the commercial sector. Internet providers and search engines protect sensitive emails and browsing histories. Merchants and credit card companies keep track of what consumers buy, and where they buy it, producing efficiencies (such as just-in-time inventory management) with only minor irritations (as when someone's personal information ends up in unwanted marketing databases). Violators are liable for civil damages, but not at the heavy hand of the Department of Health and Human Services (HHS) or the Justice Department. These systems work mainly because trust is a critical component of brand value. The same incentives apply to biomedical researchers and research institutions.[44]

Congress should convene a dedicated commission for health-care data, to promote policies that enable data-driven health research; examine the benefits and costs of medical cybersecurity; assess the viability of alternative privacy models; and consider standards of proportionality that better match the civil penalties for privacy breaches to the harms actually done. If research dollars are to be spent wisely under the Precision Medicine Initiative, it is essential that Congress adopt more pliable privacy standards, to ensure that donated genomic and phenomic data are broadly available to institutions and researchers without the fear of crippling penalties.

III. CONCLUSION

It has been predicted before, but this time it can be said with more certainty: The long-awaited rise of genomic medicine is coming. Evidence can be found in a variety of corners, from the growing efficacy of cancer cures to China's rush to seize the genomic high ground. The reverse engineering of the genetic code promises myriad benefits for every American—from lower health bills to cures for those now living on the edge of hope. World-class scientific and technical infrastructures have kept America near the front of the pack in this race for cures, but our regulatory policies have not kept pace and now constitute a strategic disadvantage. It is time to take stock of how far we have come, and to prepare our governance structures for the manifest destiny ahead.

ENDNOTES

1 Health IT spending, broadly defined, includes investments in both the technical and data infrastructures. Although there is no definitive source, various surveys suggest $100 billion may be in the ballpark. In its 2013 industry survey of industry executives, Technology Business Research Inc. estimated that health IT spending would total $34.5 billion in 2014 alone. Separately, in 2012, Insight Research estimated that spending just for telecommunications gear and services (excluding electronic health record software and data management), would total more than $100 billion during 2012-2014.

2 As of February 2011, FDA estimated that only 350 therapeutics had been approved for 7,000 rare diseases that cumulatively affect 30 million Americans. Accessed online, May 2015.

3 Ariana Eunjung Cha and Sandhya Somashekhar, "FDA Panel Debates Technique that would Create Embryos with Three Genetic Parents," *Washington Post* (online), February 25, 2015.

4 Ben Hirschler, "AstraZeneca bets on 'genetic scissors' for range of new drugs", *Reuters* (online), January 29, 2015.

5 Amy Ellis Nutt, "Bacterial DNA from Mom May be Inherited," *Washington Post* (online), Feb. 17, 2015

6 News release, "Evolutionary Surprise: Eight Percent of Human Genetic Material Comes from Viruses," University of Texas (online), January 8, 2010.

7 Eric Topol, *The Creative Destruction of Medicine*, Basic Books 2012: pp. 77-121.

8 Anonymous, "Jewels in the Junk", *The Economist* (online), June 14, 2007

9 Georges St. Laurent, Yuri Vyatkin and Philipp Kapranov, "Dark matter RNA illuminates the puzzle of genome-wide association studies", *BioMed Central* (online) June 12, 2014

10 Described online March 2015

11 Anonymous, "RNA: Really New Advances", *The Economist* (online), June 14, 2007

12 Monte Morin, "Gene Therapy Shows Promise Against HIV, Study Notes," *Washington Post* (online), February 19, 2015: A3

13 J. De Leon and D.M. Nikoloff, "Paradoxical excitation on diphenhydramine may be associated with being a CYP2D6 ultrarapid metabolizer: three case reports," *PubMed* (online), February 13, 2008. More research is needed to establish a scientifically recognized connection between CYP2D6 variation and Benadryl. However, variations in the CYP2D6 gene are associated with the toxicity of codeine, warfarin and other medications.

14 Courtney Humphries, "Shedding Light on Life", *Harvard Magazine* (online), January-February 2015.

15 National Human Genome Research Institute, "DNA Sequencing Costs", *NIH*, found online, Mar. 2015

16 Illumina marketing document (online) March 2015

17 Christina Larson, "Inside China's Genome Sequencing Factory," *MIT Technology Review* (online), February 11, 2013

18 Amazon Elastic File System storage costs were $.30 per GB per month as of May 26, 2015 (online).

19 Daniel Castro and Jordan Misra, "The Internet of Things," Center for Data Innovation, (online) November 18, 2013: pp 19-24.

20 Charles D, Gabriel M, Searcy T. "Adoption of Electronic Health Record Systems among US Non-Federal Acute Care Hospitals: 2008-2014," ONC Data Brief, April 2015 (online)

21 Michael C. Schatz and Ben Langmead, "The DNA Data Deluge," *Spectrum* (online), June 27, 2013.

22 Christopher G. Chute, Mollie Ullman Culere, et. Al, "Some experiences and opportunities for big data in translational research", *Genetics in Medicine, Vol 15 No. 10*, October 2013 (online)

23 Op. Cit, Schatz and Langmead

24 CMS Office of the Actuary, National Health Expenditure Projections 2013-2023 (online), September 17, 2014, and authors' calculations. Institute of Medicine, "Better Care at Lower Cost: The Path to Continuously Learning Health Care in America," published (online), September 2012.

25 Congressional Budget Office, "Offsetting Effects of Prescription Drug Use on Medicare's Spending for Medical Services" (online), November 2012 . CBO's rule is as follows: "…a 1 percent increase in the number of prescriptions filled by beneficiaries would cause Medicare's spending on medical services to fall by roughly one-fifth of 1 percent."

26 Marla Paul, "Persuading Doctors to Quickly Adopt New Treatments", Northwestern University (online), October 27, 2014

27 Author's calculations reflect cost of insuring a family of four in 2015 See: Kaiser-HRET (online). Waste is assumed to be 25%, but other estimates put it at 49%. See: Donald M. Berwick, MD, MPP; Andrew D. Hackbarth, "Eliminating Waste in U.S. Health Care," JAMA (online) April 11, 2012

28 Joseph A. DiMasi, slides: "Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs," Tufts Center for the Study of Drug Development, November 18, 2014. Accessed online March 2015

29 Steven M. Paul, Daniel S. Mytelka, et al., "How to improve R&D productivity: the pharmaceutical industry's grand challenge," Nature, vol. 9 (online), March 2010

30 Food and Drug Administration, accessed online March 2015

31 Op. Cit., Paul, Mytka, et al.

32 The President's Council of Advisors on Science and Technology, "Report To The President On Propelling Innovation In Drug Discovery, Development, and Evaluation," Executive Office of the President (online), September 2012

33 Op. Cit, Larson.

34 Robert D. Atkinson, Stephen J. Ezell, et al., "Leadership in Decline: Assessing U.S. International Competitiveness in Biomedical Research", The Information Technology and Innovation Foundation and United for Medical Research (online), May 2012

35 Chun-Ju Hsiao, Esther Hing and Jill Ashman "Trends in Electronic Health Record System Use Among Office-based Physicians: United States, 2007–2012", National Health Statistics Report No. 75, May 20, 2014 (online)

36 Donald M. Berwick, MD, MPP; Andrew D. Hackbarth, "Eliminating Waste in U.S. Health Care," JAMA (online) April 11, 2012

37 Allison B. McCoy et al., "Matching identifiers in electronic health records: implications for duplicate records and patient safety," BMJ Quality and Safety (online), 2013

38 Covered entities must provide requested health information within 30 days, unless the data is located off-site, in which case covered entities are given an additional 30 days. In addition, covered entities can request a one-time 30 day extension under extenuating circumstances. 45 CFR 164.524 - Access of individuals to protected health information (online). Stage 1 meaningful use criteria requires certified EHRs to provide patients that ability to download and transmit information within four days of being made available to the healthcare provider. "Eligible professional Meaningful Use Table of Contents Core and Menu Set Objectives," Centers for Medicare and Medicaid Services (online).

39 Babbage, "Computer-aided Medicine: Dr. Watson", *The Economist* (online), February 14, 2013

40 Erika Check Hayden, "U.S. Regulators try to tame 'wild west' of DNA testing," *Nature* (online) February 20, 2015.

41 45 CFR 46. The Common Rule applies uniformly across agencies, but is administered by HHS (online).

42 See: Institutional Review Board Guidebook (accessed online on May 26, 2015). IRBs are advisory bodies created to implement the Common Rule, which pre-dates HIPAA. The Guidebook was last updated in 1993.

43 Melissa Gymrek, Amy L. McGuire, David Golan, Eran Halperin, Yaniv Erlich, "Identifying Personal Genomes by Surname Inference," *Science*, Vol. 339, January 13, 2013 (online).

44 Yaniv Erlich, James B. Williams, et al., "Redefining Genomic Privacy: Trust and Empowerment", *PLOS Biology, Vol 12, Issue 11* (online), November 2014

ACKNOWLEDGMENTS

From Evolution to Revolution: Building the 21st Century Genomic Infrastructure is a white paper written by Health IT Now and the Center for Data Innovation based on ideas discussed by participants at the conference, *"Data-Driven Medicine in the Age of Genomics,"* held in Washington, D.C. in December 2014. The paper chronicles recent developments in the field of genomics and precision medicine, examines the institutional barriers to further progress in the field, and recommends regulatory and policy approaches designed to improve research productivity and expedite cures.

Both organizations would like to thank Paul Hewitt, the leading author of this white paper, as well as those conference participants who took the time to read and comment on the paper's numerous drafts, specifically Dr. Toby Bloom from the New York Genome Center and Dr. Josh Peterson from Vanderbilt University School of Medicine. Furthermore, we would like to thank Oracle, IBM, Intel and PhRMA for their support of the conference.

ABOUT HEALTH IT NOW

Health IT Now is a broad based coalition of patient groups, provider organizations, employers and payers that support incentives to deploy health information technology to improve quality, outcomes, and patient safety and to lower costs. For more information, please go to www.healthitnow.org. Follow us on Twitter at @HealthITNow

ABOUT CENTER FOR DATA INNOVATION

The Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. Based in Washington, DC, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. Learn more at www.datainnovation.org. Follow us on Twitter at @DataInnovation