HARVARD
School of Engineering
and Applied Sciences

The Genetic Reformation

Rethinking Autonomy and Data Privacy | by Simson L. Garfinkel

***On April 14, 2011**, experts in medical ethics, law, public policy, research, and entrepreneurship gathered in Cambridge for a symposium on "**Privacy, Autonomy, and Personal Genetic Information in the Digital Age**."*

Co-hosted by Cherry A. Murray, Dean of the Harvard School of Engineering and Applied Sciences (SEAS), and Leslie Berlowitz, President of the American Academy of Arts and Sciences, two panel discussions examined the "promise and perils" of creating digital repositories for genetic records and considered the policy implications of an individual's right to access, control, and interpret his or her own genetic data.

The interdisciplinary event, drawing on expertise across the Harvard campus and from around the world, was held in conjunction with a Stated Meeting of the American Academy and regional meetings of the National Academy of Engineering and the Institute of Medicine.

This edition of **Topics** *explores in greater depth some of the issues raised at that "Triple Academies" event.*

*The feature article (below) by **Simson L. Garfinkel**, Associate Professor at the Naval Postgraduate School, a noted technology writer and former postdoctoral researcher in computer science at SEAS, is intended as an engaging starting point for discussion. (The views expressed are not necessarily those of Harvard or SEAS, but rather are meant to provoke further debate and exploration.)*

Roughly nine months before you were born, your biological mother and father wrote a book. They filled that book with their hopes, dreams and plans for your life. They wrote about the adversities you might have to overcome. And they inscribed your family's secrets—long forgotten infidelities, insanity, and distant cousins who might be monsters.

Now imagine that the book was locked away and lost—only to be found decades later by the trustee of some scientific organization. That trustee may hold the keys to your future, for you are the very person that your parents wrote about. But you've also changed—you are much more today than you were when it was written. Do you have a right to decide who reads that book, once it's found?

And if so, should you read it?

Various forms of genetic testing have been available for decades. One of the first widespread testing efforts started in 1969 among Ashkenazi Jews to see if they were carriers of Tay-Sachs, a recessive genetic disease. Because it is recessive, Tay-Sachs has a 25% chance of striking the child of two carriers. But until the testing effort, no one knew who the carriers were. Because the disease is always fatal, the testing effort had but one achievable goal: prevent the conception (or at least the birth) of children who would surely die. The success of this program was one of the first great achievements of genetic testing.

Today genetic testing is widespread. New York state, for example, mandates the screening of newborns for 40 different diseases and disorders. Most of these diseases impact fewer than 1 in 10,000 newborns and can be readily treated with a special diet. Because of the testing, many children are able to lead healthy lives—children who otherwise would have died.

For example, 1 in 19,000 children are born with phenylketonuria (PKU), a disease characterized by an inability to metabolize phenylalanine, a commonly occurring amino acid. People with PKU who "diet for life" (by avoiding milk, eggs, the artificial sweetener aspartame, and other foods) are able to lead normal lives. Those who don't, suffer delayed development, mental retardation, and a variety of other problems. Another victory for large-scale genetic testing.

> *Just as Bibles translated into the vernacular helped power the Protestant Reformation, direct-to-consumer genetic testing is opening the door to a genetic reformation.*

Genetic testing also gave rise to a new professional class—the genetic counselor. Like the priests of old, these people were trained in the intricacies of an unfamiliar language—although this language was the As, Cs, Gs and Ts of the genetic code, rather than the *tempus nascendi, et tempus moriendi* ["a time to be born and a time to die"—Ecclesiastes 3:2] of the Latin Vulgate Bible. But like priests, genetic counselors were intermediaries, standing between the laity and a higher authority. And they were needed, because until the 1990s, most Americans not only lacked the ability to interpret their test results; they didn't even have legal access to their own medical records.

Now, for the first time in history, anyone on the planet who has a few hundred dollars (and is willing to spit into a tube for 15 minutes) can get vast amounts of genetic information with no intermediary whatsoever. In a few years, you'll be able to get your entire genomic sequence for less than $1,000. (You can order it today for $4,995 from Knome, Inc., a life sciences company in Cambridge, Mass.) Last year, an advisory panel told the U.S. Department of Defense that it needed to start planning for the advent of the "$100 genome"—and with it the possibility that American soldiers might be covertly tested by the enemy.

Just as Bibles translated into the vernacular helped power the Protestant Reformation, direct-to-consumer (DTC) genetic testing is opening the door to a genetic reformation. That reformation will fundamentally change our notions of ourselves, our place in the world, and our human potential. And anyone who takes the plunge will find that this genetic data brings them into a new world—one in which traditional authorities have less influence and individuals have less privacy and greater risk. And yet, the actual scientific payoff is still largely unknown.

The Personal Genetics Revolution, Right In Your Web Browser

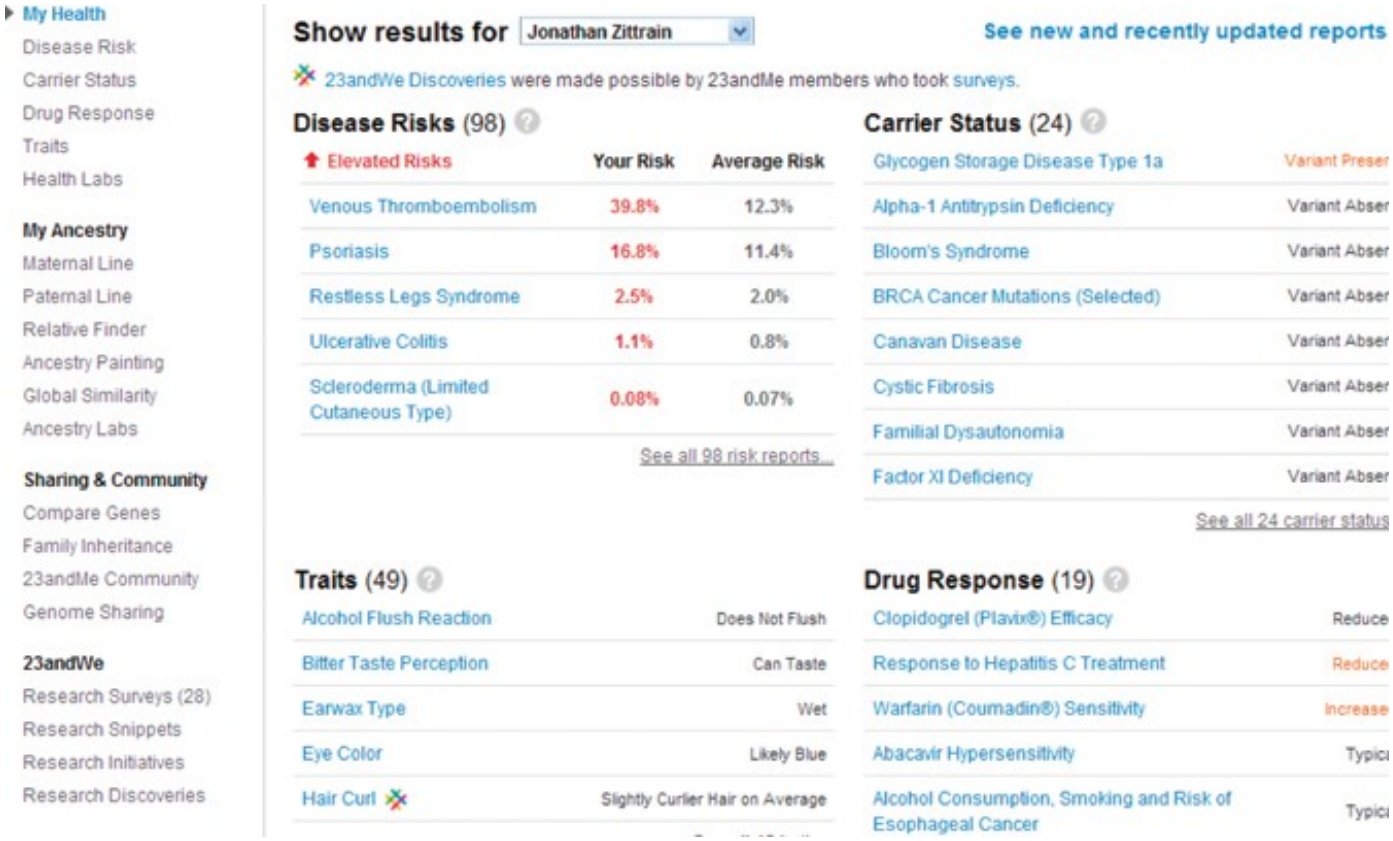
Several websites, such as deCODEme.com and Navigenics.com, offer a variety of genetic tests directly to the consumer for between $500 and $1,500. But the poster child for the genetics reformation is unquestionably 23andMe.com, a Google-backed Silicon Valley start-up that offers broad-spectrum genetic testing for about $100 (provided you sign up for a $5/month monitoring service).

Signing up is easy. Just go to a website; accept the frightening consent statement; type in a credit card number; and a few days later 23andMe's DNA collection kit will appear in your mailbox. The "kit" is really nothing more than a tube with a fancy lock and some preservatives. Avoid eating for 30 minutes; fill the tube with spit; snap on the cap; and send it back. A few weeks later you'll be able to browse your risk factors for more than 90 diseases and traits on the company's website.

Read a personal account of the author's experience with DTC genetic testing here.

But genetic analysis is only the beginning of what 23andMe does with your data. Recall that each of your parents contributed half of the words in your genetic book. This means that if you and your long-lost sister (or half-sister) both sign up for 23andMe, the company's computer can match the two of you by the fact that 50% of your genetic material is in common. And here's where things get sticky. 23andMe can also determine to a high probability that you and that woman across town have the same father but different mothers; it will even let you contact each other through the company's website.

You didn't know you had a half-sister? Oh my! That's why 23andMe's consent statement reads, in part, "You may learn information about yourself that you do not anticipate," and "Once you obtain your genetic information, the knowledge is irrevocable."

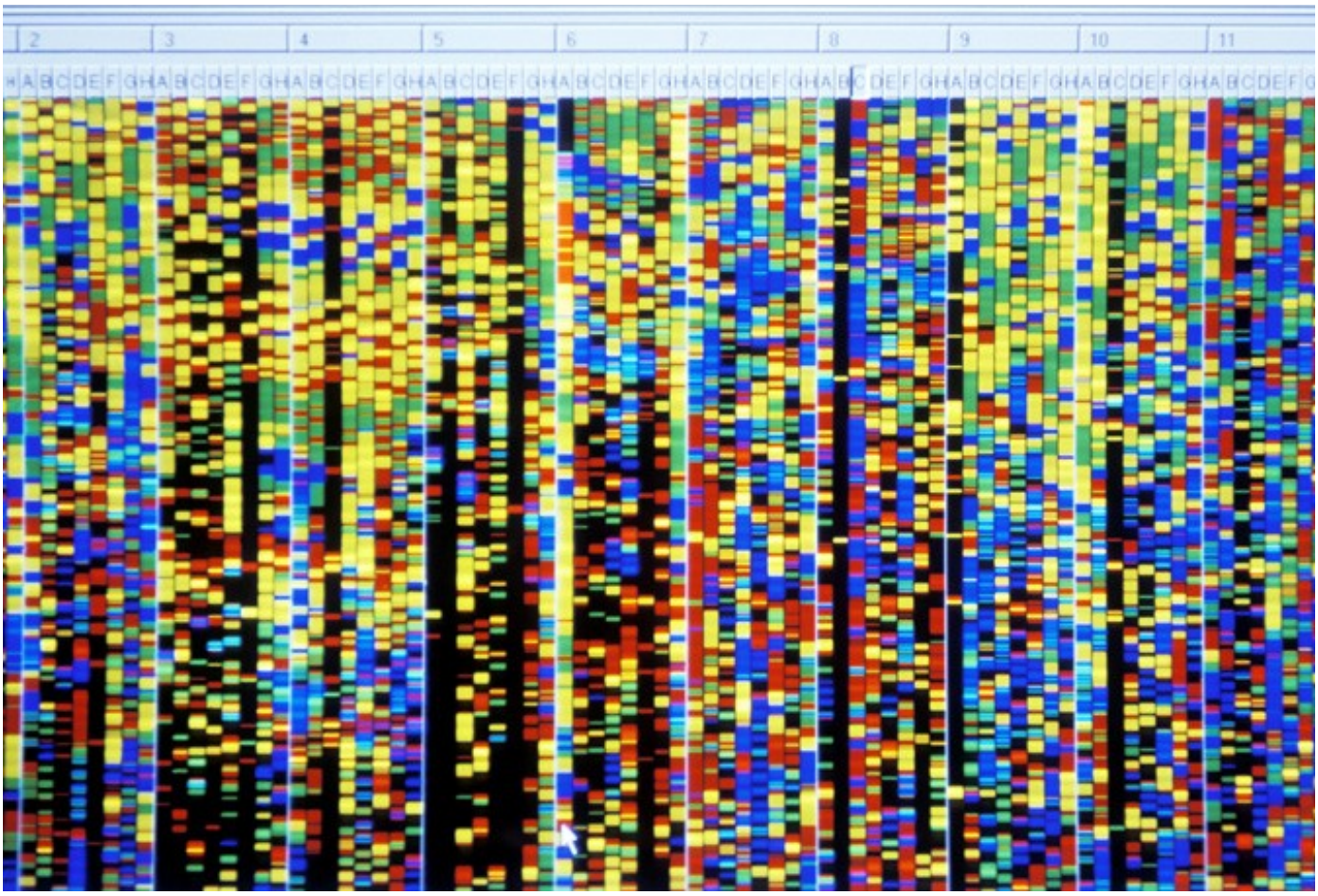


Jonathan Zittrain (M.P.A. '95, J.D. '95), Professor of Computer Science and Law at SEAS and Harvard Law School, submitted a saliva sample to a personal genetic testing service to see what it was all about. Now, when he logs into 23andMe.com, the website informs him of his risks and traits, including his earwax type—"wet." (Image courtesy of Jonathan Zittrain.)

For investors, another significant part of 23andMe is likely to be the company's research arm—23andWe. Human genetics is vastly more complicated than the simple Mendelian genetics that most of us learned in high school. Disease susceptibility, drug response—even "simple" things like hair color and curl—actually result from the interaction of multiple genes along with environmental challenges and other factors that modern biology is only just beginning to understand. So in addition to testing the genotype of each subscriber, 23andMe invites them to participate in research studies by answering detailed questionnaires about their phenotype—that is, their medical history, traits, and morphology, such as the shape of their noses and the character of their ear wax. All of these data are mined in an effort to draw correlations between genetic sequence, medical conditions, and medical outcomes.

Unlike the 40 tests mandated by the state of New York, 23andMe isn't performing traditional genetic tests. When children in New York are tested for PKU, they are tested to see if an enzyme called phenylalanine hydroxylase is present and properly functioning—in other words, that the individual's genes are correctly producing that enzyme. 23andMe doesn't assess the presence or absence of given enzymes or proteins. Instead, it screens for what are called single nucleotide polymorphisms —SNPs (pronounced "snips") in the language of the new biology. A SNP is a variant spelling in the genome written by our biological parents.

If you think of each person who's alive today as having a book in humanity's genetic library, it turns out that those 7 billion books are remarkably similar. In the vast majority of cases, these books have 46 chapters, representing 23 chromosomes from each parent. Each book is thought to have 20,000 to 25,000 pages—one page for each protein-coding gene. Though there are roughly 3 billion base pairs in total—3 billion As, Cs, Gs, and Ts—most of these don't code for genes, but seem to be associated with some form of regulation. Yet, in all of this writing that's so important in determining our destiny, there are only about 10 million places where one person's "T" might be another person's "G," or where another person might have a few extra Ts—as if his or her mother or father momentarily stuttered. And according to the 23andMe website, the company now measures roughly 1 million of these 10 million SNPs from the 23 chromosomes, as well as a few thousand SNPs from the mitochondrial DNA.



Advances in bioinformatics—the application of computer science to biology and medicine—have been crucial for the field of genomics. Fast and accurate assembly of complete genomes would not be possible without sophisticated sequencing algorithms, modeling techniques, and (as shown above) data visualization tools. In this computer readout, each color corresponds to a nucleotide (A, C, T, or G) detected in a genomic sequence. (Photo by Patrick Landmann / Photo Researchers, Inc.)

Although reading the SNPs is a highly precise technique for measuring a person's genetic profile, for many genetic diseases it is far more accurate to simply measure the presence or absence of a functioning enzyme. That's because, in the case of PKU, there are literally hundreds of different genetic mutations that might cause a child to have an absent or poorly functioning phenylalanine hydroxylase enzyme. While some of these mutations are known and reported in the medical literature, others aren't.

For the cases where the mutation is known, the 23andMe website tells subscribers if their SNPs match the literature. The website will even cite the study, allowing consumers-turned-scientists to examine the literature for themselves.

For cases in which the link between the genotype (the specifics of the genetic plan) and phenotype (the expressed, physical characteristics) is unknown, 23andMe hopes that those surveys will help scientists to draw correlations between various SNPs and the prevalence of various diseases. For example, if 75% of 23andMe's subscribers who have a "T" in a particular position tend to get some disease by age 30, and most people who have an "A" there do not, then that "T" might be associated with a 75% chance of contracting the disease. Then again, it might not.

To put this into actual genetic terms, consider leucine-rich repeat kinase 2, an enzyme coded by the *LRRK2* gene, one of the 1,370 genes on chromosome 12. A specific mutation in this gene called *rs34637584(A)* is associated with a significantly higher chance of contracting Parkinson's disease. Years ago, finding out what such a mutation meant would have required a master's degree and access to a medical library. Today you can type the mutation into Google and be directed to a page on SNPedia, an open-source wiki devoted to collecting such information and making it public. The wiki says, in part:

One copy of a[n] rs34637584(A) *allele is sufficient to greatly increase one's risk for Parkinson's disease. … Overall, the risk of Parkinson's disease for*

a person who inherits a[n] rs34637584(A) *allele is 28% at age 59, 51% at 69, and 74% at 79, according to the International LRRK2 Consortium.*

Unfortunately, the wiki misstates the evidence.

As all students learn during their first statistics course, *correlation is not causation*! The *rs34637584(A)* allele may be more likely to be present in people who have Parkinson's disease, but we don't know if it is the cause—the SNP might work in concert with another gene, or with an environmental agent, or it may be an innocent genetic bystander.

DTC's Organized Opposition

It's these uncertainties, in part, that have caused organizations such as the American Medical Association, the National Society of Genetic Counselors, and the American Society of Human Genetics to call for significant regulation of DTC companies. In February 2011, the American Medical Association's executive vice president, Dr. Michael D. Maves, wrote to the Food and Drug Administration, urging that tests "with the highest risk of harming consumers if misinterpreted have the strictest regulatory requirements," and recommending that companies like deCODE Genetics and 23andMe be legally required to report these test results to a customer's physician or genetic counselor, and not directly to the consumer.

Indeed, the state of New York already prohibits companies from offering direct-to-consumer genetic tests. As a result, when someone in Manhattan wants to be tested by 23andMe, the company requires that the specimen be mailed from outside New York—for example, by taking a 10-minute subway ride to New Jersey and dropping the package into a Hoboken mailbox.

Doctors and genetic counselors who want legislative prohibitions on DTC testing are clearly acting in their own economic interest: each consumer who bypasses today's inefficient healthcare system and goes directly to these companies is saving hundreds, if not thousands, of dollars. Wiki-based counseling is free.

On the other hand, it's easy to take the genetic priesthood's claims at face value: this is powerful information and easily misinterpreted. There are documented cases of people committing suicide after learning that they were carriers for Huntington's disease—and those people received counseling in a clinical environment. We have no idea how much damage might be done in the coming years by the casual release of such sensitive medical information.

> *Unlike a stolen credit card number, a genome can't be changed if it is inadvertently given to criminals. Is that a risk? We just don't know.*

Another danger is that this information might not be adequately protected. 23andMe allows its customers to download their entire genetic data set: other websites invite you to upload it for a third-party analysis. The problem here, of course, is that unlike a stolen credit card number, a genome can't be changed if it is inadvertently given to criminals. Is that a risk? We just don't know.

There's another potential problem with DTC genetic testing: is it accurate?

"We are concerned about analytic validity," says Dr. Michele Caggana, Section Head for Genetic Testing for the New York State Clinical Laboratory Reference System. "If you order the test 10 times, do you get the same results 10 times?"

The Government Accountability Office (GAO), the watchdog agency of the U.S. Congress, has twice reviewed DTC testing firms and found troubling inconsistencies, reporting in 2006 that the firms made "medically unproven disease predictions." A 2010 GAO report—"Direct-To-Consumer Genetic Tests: Misleading Test Results Are Further Complicated by Deceptive Marketing and Other Questionable Practices"—was even more damning. It constituted an outright attack on the industry, accusing it of inconsistent test results, incorrect information delivered by telephone consultants, and the use of genetic information to scare customers into purchasing expensive vitamin supplements.

The GAO did not release the names of the companies that it investigated, but it did refer them to the Food and Drug Administration and the Federal Trade Commission "for appropriate action."

In a high-profile June 2010 incident, 23andMe mixed up the samples in a 96-well plate and sent incorrect DNA results to 96 of its customers. Whoops! One family, which had tested parents as well as children, was taken aback—they thought that their son might

have been accidentally swapped at the hospital when he was born (apparently it had happened a month before at the same hospital to another pair of babies). Another company, deCODE Genetics, had a similar problem in August 2009.

Risky Business

For many contemplating DTC tests, issues of cost and accuracy are less important than the potential damage that might come from taking the test—not just the way that genetic information might change one's sense of self, but the real potential for genetic discrimination.

There's a long history of using genetics and pseudo-genetics to justify discrimination against individuals and racial groups—and you don't need to go back to the Second World War for examples. Since the 1980s, the Council for Responsible Genetics has documented more than 500 cases in which apparently healthy individuals have been "barred from employment or lost their health and life insurance based on an apparent or perceived genetic abnormality," according to CRG's project on Genetic Testing, Privacy, and Discrimination. Fears of genetic discrimination were also taken to the big screen in the 1997 movie *Gattaca*.

In 2008, Congress passed the Genetic Information Nondiscrimination Act (GINA), which prohibits the use of genetic information for determining health insurance rates or employment. But GINA does allow genetic tests to be used for setting rates on long-term care insurance and life insurance. This means that women who are tested and found to carry harmful mutations of the *BRCA1* or *BRCA2* genes—that is, women who have more than a 50% chance of developing breast or ovarian cancer—can't be denied a job or health insurance, but they can be denied life insurance.

Is genetic discrimination a compelling risk? Dr. Philip Reilly, who spent years caring for institutionalized individuals with genetic disorders and now, at Third Rock Ventures in Boston, invests in companies that are trying to treat them, insists that it's not.

"We have a 40-year history of gathering, storing, and protecting" genetic information, Reilly says. "There is virtually no evidence that anyone has suffered an economic harm from newborn screening. It's appropriate to think about the [potential for abuse], but it's outweighed by the benefits."

The trouble with this argument is that newborns haven't been screened for their risk of contracting Parkinson's or Alzheimer's diseases later in life—two diseases that have profound financial impact on those offering life or long-term care insurance.

"To what extent can one actually know the consequence of

releasing that data?" asks Latanya Sweeney (A.L.B. '95) (*pictured*), a Visiting Professor of Computer Science at Harvard's School of Engineering and Applied Sciences (SEAS). Sweeney contends that we simply don't know the ways that this information could be abused.

Even if there are no direct harms, many feel that it is a violation of personal privacy to release even anonymous genetic information without consent. And that's exactly what happened in Texas between 2002 and 2009, when 8,350 of the 5.3 million samples collected during the course of its newborn screening program were released to 27 separate research programs by scientists around the United States.

In March 2009, shortly after news of the medical research was made public, Geoffrey Courtney of San Antonio and four other parents filed suit against the Texas Health Department and Texas A&M University, alleging that the state's retention of the blood spots and their use in research and federal investigations constituted unlawful search and seizure and violated their privacy rights. The suit was settled out of court, but in response, the Texas legislature passed a law specifically authorizing this use of the blood spots —provided that the parents were first
allowed to opt out of the collection by signing a form. In early 2010, Texas incinerated the 5.3 million blood spots that had been collected prior to the passage of the law.

But there are problems even with anonymous genetic samples. When such "de-identified" samples are released for research, the anonymization is really just a legal fiction. Like a picture of a face or a fingerprint, genes carry information that is hugely personal. With enough ancillary information—for example, a large database that happens to have DNA SNPs from other family members —re-identification is quite possible.

"It might be possible 10 years from now to actually anonymize genetic data, but it's not possible now because we don't know what it is we need," says Sweeney, who has worked on techniques for re-identifying medical records and other kinds of information for more than decade. "If I need all of it, then I can't really de-identify it, because it's you."

Another thing that can't be de-identified is familial relationships. With a database of SNPs from thousands of people, it's now fairly straightforward to identify who's related to whom. And because the DNA molecule is stable over a long period of time—you can recover DNA from corpses that are thousands of years old—blood or body specimens from people who died in the 1950s could easily be used to learn sensitive information about people living today. This creates a paradox under U.S. law, since the dead legally have no privacy rights.

The ready availability of genetic information, made possible by direct-to-consumer genetic tests, thus creates fundamental challenges to our notions of privacy, autonomy, and consent. Given the shared nature of genetic information, it may be a fundamental misconception to view this data as "private." Indeed, Dr. George M. Church, Professor of Genetics at Harvard Medical School and Director of the Center for Computational Genetics, has published a document arguing that there are so many ways that genomic confidentiality might be compromised, that "guarantees of genome anonymity" are simply unrealistic.

Dr. George Church, Ph.D. '84, Professor of Genetics at Harvard Medical School, created the Personal Genome Project in 2005 with the hope of gathering 100,000 individuals' genomic sequences and medical histories for scientific research. Almost 30 years ago, while earning his Ph.D. at Harvard in biochemistry and molecular biology, he helped to develop the first direct genomic sequencing method. Today, he combines genomics with epigenetics and developmental biology to study stem cells. (Photo by Volker Steger / Photo Researchers, Inc.)

Making Personal Genomes Public

There's a lot we don't know about the impact of personal genetic information on individuals, families, and society as a whole. Would learning that you have an 80% chance of dying from heart disease by age 40 lead you to embrace a healthier lifestyle in the hope of beating the odds—or simply justify continued pigging out on those high-fat foods, in the belief that you can't change your genetic destiny? We just don't know.

Likewise, we don't know how many complete genomes are needed to make fundamental discoveries. The *LRRK2* Parkinson's study mentioned in the SNPedia entry was based on a study of just 1,045 people from 133 families—and for those people, the study confined itself to the *LRRK2* gene. Perhaps better science would have been possible if more people had been studied, and if the entire genome for those symptomatic individuals had been made available.

Questions like these are at the heart of the Personal Genome Project, a multi-year research effort headed by Prof. Church at Harvard Medical School. This federally funded project is seeking volunteers who will consent to having their entire genome sequenced and made freely available on the Internet with the goal of aiding scientific discovery. But the project also hopes to study the volunteers apart from their genomes, exploring the impact of genetic information and education on them and their families.

Ultimately, the PGP seeks to collect and publish the genomes for 100,000 individuals. The first 10 individuals are also sharing their detailed medical records and other highly personal information. These so-called "PGP-10" include Church; venture capitalist Esther Dyson; the CEOs of several genomic-based healthcare organizations; and research scientists from Harvard and Duke.

For the CEOs and scientists involved there was a clear benefit in making their genomes available: they hope to profit from the availability of scientific research data. But what about the other 99,990 people that Church wants to recruit? For those, the project hopes to attract individuals with a combination of personal curiosity and "genetic altruism"—people who feel that, by sharing their genome, they can help make the world a better place.

Just as the publishing and discussion of Martin Luther's 95 Theses powered the Protestant Reformation, it's almost certain that the growing availability of genetic information from DTC genetic testing and online genomic sharing will result in profound changes, producing a genetic reformation. Our challenge as a society is to guide this reformation in such a way that it maximizes the benefits of the information while minimizing the potential for harm.

Harvard School of Engineering and Applied Sciences
29 Oxford Street, Cambridge, MA 02138, USA