



Sign In | RSS Feeds

All Wired

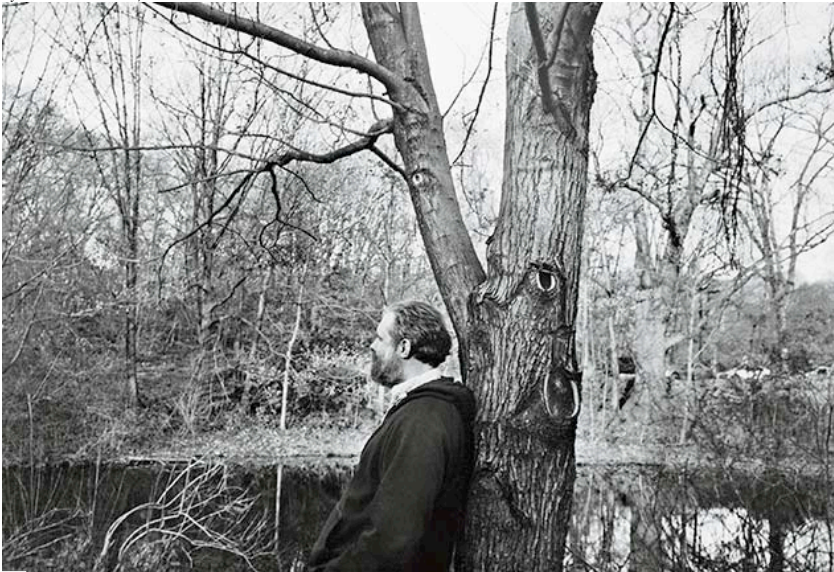
GO

WIRED MAGAZINE: 16.08

Med-Tech : Stem Cells

How the Personal Genome Project Could Unlock the Mysteries of Life

By Thomas Goetz 07.26.08



George Church is sequencing the DNA of 100,000 volunteers.

Photo: Lloyd Ziff

George Church is dyslexic, narcoleptic, and a vegan. He is married with one daughter, weighs about 210 pounds, and has worn a pioneer-style bushy beard for decades. He has elevated levels of creatine kinase in his blood, the consequence of a heart attack. He enjoys waterskiing, photography, rock climbing, and singing in his church choir. His mother's maiden name is Strong. He was born on August 28, 1954.

If this all seems like too much information, well, blame Church himself. As the director of the Lipper Center for Computational Genetics at Harvard Medical School, he has a thing about openness, and this information (and plenty more, down to his signature) is posted online at arep.med.harvard.edu/gmc/pers.html. By putting it out there for everyone to see, Church isn't just baiting identity thieves. He's hoping to demonstrate that all this personal information — even though we consider it private and somehow sacred — is actually fairly meaningless, little more than trivia. "The average person shouldn't be interested in this stuff," he says. "It's a philosophical exercise in what identity is and why we should care about that."

As Church sees it, the only real utility to his personal information is as data that reflects his phenotype — his physical traits and characteristics. If your genome is the blueprint of your genetic potential written across 6 billion base pairs of DNA, your phenome is the resulting edifice, how you actually turn out after the environment has had its say, influencing which genes get expressed and which traits repressed. Imagine that we could collect complete sets of data — genotype and phenotype — for a whole population. You would very quickly begin to see meaningful and powerful correlations between particular genetic sequences and particular physical characteristics, from height and hair color to disease risk and personality.

Church has done more than imagine such an undertaking; he has launched it: The Personal Genome Project, an effort to make those correlations on an unprecedented scale, began last year with 10 volunteers and will soon expand to 100,000 participants. It will generate a massive database of genomes, phenomes, and even some *omes* in between. The first step is to sequence 1 percent of each volunteer's genome, focusing on the so-called exome — the protein-coding regions that, Church suspects, do 90 percent of the work in our DNA. It's a long way from sequencing all 6 billion nucleotides — the As, Ts, Gs, and Cs — of the human genome, but even so, cataloging 60 million bits multiplied by 100,000 individuals is an audacious goal.

The PGP stands as the tent pole of what Church calls his "year of convergence," the moment when his 30 years as a geneticist, a technologist, and a synthetic biologist all come together. The project is a proof of concept for the Polonator G.007, the genetic-sequencing instrument developed in Church's lab that hit the market this spring. And the PGP will also put Church's expertise in synthetic biology to use, reverse engineering volunteers' skin cells into stem cells that could help diagnose and treat disease. If the convergence comes off as planned, the PGP will bring personal genomics to fruition and our genomes will unfold before us like road maps: We will peruse our DNA like we plan a trip, scanning it for possible detours (a predisposition for disease) or historical markers (a compelling ancestry).

Bringing the genome into the light, Church says, is the great project of our day. "We need to inspire our current youth in a way that outer space exploration inspired us in 1960," he says. "We're seeing signs that knowing about our inner space is very compelling."

To Church, who built his first computer at age 9 and taught himself three programming languages by 15, all of this is unfolding according to the same laws of exponential progress that have propelled digital technologies, from computer memory to the Internet itself, over the past 40 years: Moore's law for circuits and Metcalfe's law for networks. These principles are now at play in genetics, he argues, particularly in DNA sequencing and DNA synthesis.

Exponentials don't just happen. In Church's work, they proceed from two axioms. The first is automation, the idea that by automating human tasks, letting a computer or a machine replicate a manual process, technology becomes faster, easier to use, and more popular. The second is openness, the notion that sharing technologies by distributing them as widely as possible with minimal restrictions on use encourages both the adoption and the impact of a technology.

Inside the Personal Genome Project

The project will turn information from 100,000 subjects into a huge database thath can reveal the connections between our genes and our physical selves. Here's how. — Thomas Goetz



Email Article

Full Page
Comments

A small image of a Wired magazine cover, showing the title "WIRED" and the subtitle "P=φ(A,B,Y)".

subscribe to WIRED

Subscribe to WIRED

Renew

Give a gift

Customer Service

SERVICES

Subscription: [Subscribe](#) | [Give a Gift](#) | [Renew](#) | [International](#) | [Questions](#) | [Change Address](#)

Quick Links: [Contact Us](#) | [Login/Register](#) | [Newsletter](#) | [RSS Feeds](#) | [Tech Jobs](#) | [Wired Mobile](#) | [FAQ](#) | [Sitemap](#)

1. Entrance Exam

Volunteers take a quiz to show genetic literacy. One question: How many chromosomes do unfertilized human egg cells contain? a) 11, b) 22, c) 23, d) 46, or e) 92? (Answer: c.) Only those with a perfect score proceed, but retests are allowed.

2. Data Collection

Volunteers sign an "open consent" form acknowledging that their information, though anonymized, will be accessible by others. They fill out their phenotype traits, listing everything from waist size to diet habits. Suitable respondents go on to the next step.

3. Sample Collection

Volunteers hit the medical center, where they are interviewed by an MD. Then a technician draws some blood, gathers a saliva sample, and takes a punch of skin. Don't worry: It hurts about as much as a bee sting.

**4. Lab Work**

The tissues are sent to a biobank, where DNA is extracted from the blood. One percent of it — the exome — is sequenced. Meanwhile, bacteria DNA is extracted from the saliva and sequenced to reveal the volunteer's microbiome.

**5. Research**

Now the fun part: Crunching the numbers. PGP scientists and other researchers start working with the data assembled from 100,000 individuals to investigate potential links between phenotypes and genotypes. The team will look for patterns and statistically significant anomalies.

**6. Sharing**

The volunteers get access to not only the raw data from their genome, but anything the research team gleans from their information. Insights — a newly discovered cancer risk, for example — are posted in a volunteer's file, which they'll be free to share with other PGP participants.

"I always tell people, your biggest problem in life is not going to be hiding your stuff so nobody steals it," Church says. "It's going to be getting anybody to ever use it. Start hiding it and that decreases the probability to almost zero."

For most of his career, Church has been known as a brilliant technologist, more behind-the-scenes tinkerer than scientific visionary. Though he was part of the group that kicked off the Human Genome Project, he's far less known than scientists like Francis Collins or J. Craig Venter, who took the stage at the end. His obscurity is due partly to his style. He talks about his accomplishments with a certain detachment that one might mistake for ambivalence. "He's not without ego; it's just a different sort of ego," says entrepreneur Esther Dyson, a friend and one of the first 10 PGP volunteers. "Everything is a subject of his intellectual curiosity, including himself."

His low profile may be the result of his tendency to get too far ahead of the curve, working a decade or two ahead of his field — so far that even the experts don't always get what he's talking about. "Lots of George's work is so advanced it's not ready to become standard," says Drew Endy, a professor of bioengineering at Stanford and cofounder with Church of Codon Devices, a synthetic-biology startup. "He's perfectly happy to spin out tons of ideas and see what might stick. It's high-throughput screening for technology and science. That's not the way most people work."

But thanks to the PGP, the Polonator, and the fact that the rest of the world is finally starting to understand what he's been talking about, Church's obscurity is coming to an end. He sits on the advisory board of more than 14 biotech companies, including personal genomics startup 23andMe and genetic testing pioneer DNA Direct. He has also cofounded four companies in the past four years: *Codon Devices*, *Knome*, *LS9*, and *Joule Biosciences*, which makes biofuels from engineered algae. *Newsweek* recently tagged him as one of the 10 Hottest Nerds ("whatever that means," Church laughs). For someone who has spent his whole career ahead of his time, he is suddenly very much a man of the moment.

Most historians would cite Prague or Paris or Berkeley as the intellectual hub of the 1960s, but for people interested in computers, there was no place so significant as Hanover, New Hampshire. There, at Dartmouth College, an experiment in time-share computing was flourishing. Developed by professors John Kemeny and Thomas Kurtz, the Dartmouth Time-Sharing System let students remotely access the power of a mainframe computer to do calculations for mathematics or science assignments or to play a simulated game of college football. It ran on an easy-to-learn, intuitive program that Kemeny and Kurtz called Basic.

In 1967, the DTSS transitioned to a more-powerful GE-635 machine and offered remote terminals to 33 secondary schools and colleges, including Phillips Academy, a prep school in nearby Andover, Massachusetts. The terminal — not much more than a teletype machine, really — sat in the basement of the school's math building, forgotten until the next fall, when a young George Church showed up for his freshman year and began asking whether there was a computer on campus. Someone pointed Church to the basement. "There wasn't even a chair in the room. I had used a typewriter before, but never a teletype. And so I just started pressing keys," Church recalls. "Eventually I hit Return, and it came back with 'What?' And so I started typing in stuff like crazy and hitting Return. And it kept coming back with 'What?' At that point, I was pretty convinced it wasn't a human, but it was actually talking in words. So I just hadn't asked the right question or given the right answer."

Soon, Church found a book on Basic. "I was just sailing," he says. He spent endless hours in that basement — he eventually borrowed a chair — and taught himself the intricacies of coding, learning to program in Basic, Lisp, and Fortran. Indeed, thinking in code came so naturally to Church that he stopped going to his classes (a habit that would later get him kicked out of graduate school at Duke) and taught the computer linear algebra instead. It turns out that learning how to write code — change it, hit Return, see what it will do — was ideal training for Church's eventual career in computational biology. "That's how we reverse engineer things like *E. coli* — you change something, and you see how it behaves," he says. "Little did I know that 30 years later, we would use almost exactly the same operations to optimize metabolic networks."

Church first hit on the power of computation to automate biology in the mid-'70s when he was in graduate school at Harvard. At the time, he was working on recombinant DNA, a then-new technique to splice a gene from one organism into another. Identifying a sequence of 80 or so base pairs of genetic code was a slow, tedious process. "You had to literally read off the bases and write them on a piece of paper, one by one," Church says. "So I wrote a sequence-reading program that would crunch it out. When the senior graduate student heard I had automated that, he said, 'What do you want to do that for? That's the only fun part.'"

By 1980, when Church's adviser, Wally Gilbert, won the Nobel Prize for DNA sequencing techniques, the process was still slow and expensive, executing one DNA strand at a time. So Church began working on one of his earlier targets for automation. His idea was to sequence several strands together by combining them into a single sample mixture. He called it multiplexing, drawing an analogy to signal multiplexing in electronics, in which more than one signal flows through a current at the same time. Church thought most of the work could even be integrated into one device rather than numerous machines.

It was a provocative idea, not just because he was substituting several human tasks for machine-driven ones, but also because he didn't make the usual false promise that technology would simplify the process. On the contrary, multiplexing would be complicated, Church maintained. But technology was up to the task.

Four years later, Church was invited to present his work on multiplexing at a small meeting in Alta, Utah. The Department of Energy had gathered about 20 scientists to mull over one question for five days: How might recent advances in genetics be used to measure an increase in genetic mutations arising from radiation exposure, as in Hiroshima? The group quickly reached the conclusion that technology circa 1984 couldn't answer that question. Meanwhile, they still had several more days in the mountains. "There were a bunch of us there who could talk about genomics as if it were an engineering exercise. And then we said, well, as a kind of booby prize, we could think of other things you could do," Church recalls, "like, say, sequencing the human genome."

Though Church was almost entirely unknown before the meeting, his presentation on multiplex sequencing methods stole the show. When he fell into a huge snow drift during a break one afternoon, one participant worried that the future of sequencing had disappeared with him.

That Alta brainstorm would become the **Human Genome Project** — the effort, adopted by the National Institutes of Health, to sequence one human genome for \$3 billion within 15 years. However audacious the HGP seemed, Church was disappointed by it almost from the start. "We could have said our goal was to get everybody's genome for some affordable price," he says, "and one genome would be a milestone" on the way toward that goal.

The HGP also played it safe with its choice of technology. Despite the promise of Church's multiplexing system, the HGP instead used a more established instrument manufactured by Applied Biosystems, based on a technique developed by biochemist Frederick Sanger. As Church saw it, this meant that the project had failed to put its \$3 billion toward improving the state of the art. Even worse, the HGP consumed so many of the resources available to the field of genetics that it effectively locked that state of the art into 1980s technology.

The result was nearly two decades of inertia. It wasn't until 2005, when the Human Genome Project was complete and new goals were put forth, that Church finally perfected the multiplexing approach he had presented 20 years earlier at Alta. In a paper published in *Science*, Church demonstrated a technique that could analyze millions of sequences in one run (Sanger's method could handle just 96 strands of DNA at a time). And Church's method not only accelerated the process, it made it far cheaper, too, elegantly demonstrating the power of automation to drive exponential advances and bring down costs. Church's approach, and a competing innovation developed by **454 Life Sciences** that same year, inaugurated the second generation of sequencing, now in full swing.

In the past three years, more companies have joined the marketplace with their own instruments, all of them driving toward the same goal: speeding up the process of sequencing DNA and cutting the cost. Most of the second-generation machines are priced at around \$500,000. This spring, Church's lab undercut them all with the **Polonator G.007** — offered at the low, low price of \$150,000. The instrument, designed and fine-tuned by Church and his team, is manufactured and sold by Danaher, an \$11 billion scientific-equipment company. The Polonator is already sequencing DNA from the first 10 PGP volunteers. What's more, both the software and hardware in the Polonator are open source. In other words, any competitor is free to buy a Polonator for \$150,000 and copy it. The result, Church hopes, will be akin to how IBM's open-architecture approach in the early '80s fueled the PC revolution.

In the sequencing game, though, the cost of the machine is only half the equation. The more telling expense is the operating cost, particularly the cost of sequencing entire human genomes. Executives at 454 estimate that their latest machine can pull off a whole genome sequence for \$200,000. Applied Biosystems claims its instrument has completed a genome for just \$60,000. Church maintains that, while the Polonator isn't up to whole-genome reads, it is clocking in at about one-third the cost of Applied Biosystems' estimate. A whole sequence from [Knome](#), the retail genomics firm cofounded by Church, goes for \$350,000. (It's worth noting that these figures are only roughly comparable, since each company uses slightly different quality measures and specifications.)

As these numbers continue to drop, the mythical \$1,000 genome comes ever closer. Sequencing a human genome for \$1,000 is the somewhat arbitrary benchmark for true personalized genomics — when the science could become a component of standard medical care. An important catalyst in achieving that point is the [Archon X Prize](#) for Genomics, which is offering \$10 million to the team that can sequence 100 complete genomes in 10 days for less than \$10,000 each. As of June, seven teams, including Church's lab, had entered the competition. Church, who served for a time on the advisory board of the contest, says that the prize will drive costs down further and help publicize the potential of personalized whole-genome sequencing.

That's important because Church hopes the Polonator and other next-generation instruments will inspire a new generation of smaller labs to begin work in personal genomics, as well as other genetic sciences. Already, the onslaught of technology has jump-started new projects, like sequencing part of the Neanderthal genome, examining extremophile microbes in old California iron mines, and studying the regenerative properties of the salamander. In medicine, cheaper sequencing has enabled research into drug-resistant tuberculosis; the genetics of breast, lung, and other cancers; and the DNA architecture of schizophrenics.

But if the Polonator is going to lead that charge, it has to work — and work on a massive scale. And that means passing a major test: successfully sequencing the 100,000 exomes in the PGP.



Photo: Lloyd Ziff

All of us know our height, weight, and eye color. Fewer of us know our arm span or resting blood pressure. But who among us knows the direction of our hair whorls or the Gell-Combs type of our allergies? This is the level of detail that the PGP requires the 100,000 volunteers to reveal about themselves, a list staggering in its exhaustiveness. The PGP will tally head circumferences, injuries, chin clefts and cheek dimples, whether volunteers can roll their tongues or hyperflex their joints, whether they dislike hot climates or are hot tempered, if they've often been exposed to power lines or wood dust or diesel exhaust or textile fibers. The project questionnaire asks how many meals they eat a day and whether they prefer their food fried, broiled, or barbecued. It even demands to know how much television they watch. And, of course, PGP volunteers will hand over most aspects of their medical history, from vaccines to prescriptions.

This phenotype data will be integrated with a volunteer's genomic information, then combined with statistics from all the other subjects to create a potent database ripe for interrogation. In contrast to the heavy lifting that genetic research requires now — each study starts from scratch with a new hypothesis and a fresh crop of subjects, consent forms, and tissue samples — the PGP will automate the research process. Scientists will simply choose a category of phenotype and a possible genetic correlation, and statistically significant associations should flow out of the data like honey from a hive. A genetic predisposition for colon cancer, for instance, might be found to lead to disease only in connection with a diet high in barbecued foods, or a certain form of heart disease might be associated with a particular gene and exposure to a particular virus. Genomic discovery won't be a research problem anymore. It'll be a search function. (This helps explain why Google, among others, has donated to the project).

The process began last year, and each of the first 10 volunteers has a background in medicine or genetics. They include John Halamka, CIO of Harvard Medical School and a physician; Rosalynn Gill, chief science officer at Sciona (a personalized genetics nutrition company); and Steven Pinker, the noted psychologist and author. The other 99,990 participants won't be expected to be so elite, though they will have to pass a genetics-literacy quiz to demonstrate informed consent. The general selection process, which starts with registration at [personalgenomes.org](#), is scheduled to begin later this year.

Besides offering up their genomes, subjects will have to part with some spit and a bit of skin. The saliva contains their microbiome — the trillions of microbes that exist, mostly symbiotically, on and in our bodies. If phenotype is a combination of genotype plus environment, the microbiome is the first wash of that environment over our bodies. By measuring some fraction of it, the PGP should offer a first look at how the genome-to-microbiome-to-phenome chain plays out.

The skin sample goes into storage, creating what would be one of the world's largest biobanks. Members of Church's lab have devised a way to automate turning the skin cells into stem cells, and they hope to publish the technique later this year. (Similar work has been done at the University of Wisconsin and Kyoto University.) By reprogramming the skin cells using synthetically engineered adenoviruses, Church's team can transform the skin cells into many sorts of tissue — lungs, liver, heart. These tissues could be used as a diagnostic baseline to detect predisposition for various diseases. What's more, the reprogrammed cells could be used to treat disease, replacing damaged or failing tissue. It's an intriguing hint of how Church's work with synthetic biology complements genomic sequencing.

If the PGP were simply an exercise in breaking down 100,000 individuals into data streams, it would be ambitious enough. But the project takes one further, truly radical step: In accordance with Church's principle of openness, all the material will be accessible to any researcher (or lurker) who wants to plunder thousands of details from people's lives. Even the tissue banks will be largely accessible. After Church's lab transforms the skin into stem cells, those new cell lines — which have been in notoriously short supply despite their scientific promise — will be open to outside researchers. This is a significant divergence from most biobanks, which typically guard their materials like holy relics and severely restrict access.

For the PGP volunteers, this means they will have to sign on to a principle Church calls open consent, which acknowledges that, even though subjects' names will be removed to make the data anonymous, there's no promise of absolute confidentiality. As Church sees it, any guarantee of privacy is false; there is no way to ensure that a bad actor won't tap into a system and, once there, manage to extract bits of personal information. After all, even de-identified data is subject to misuse: Latanya Sweeney, a computer scientist at Carnegie Mellon University, demonstrated the ease of "re-identification" by cross-referencing anonymized health-insurance records with voter registration rolls. (She found former Massachusetts governor William Weld's medical files by cross-referencing his birth date, zip code, and sex.)

To Church, open consent isn't just a philosophical consideration; it's also a practical one. If the PGP were locked down, it would be far less valuable as a data source for research — and the pace of research would accordingly be much slower. By making the information open and available, Church hopes to draw curious scientists to the data to pursue their own questions and reach their own insights. The potential fields of inquiry range from medicine to genealogy, forensics, and general biology.

And the openness doesn't serve just researchers alone. PGP members will be seen as not only subjects, but as participants. So, for instance, if a researcher uses a volunteer's information to establish a link between some genetic sequence and a risk of disease, the volunteer would have that information communicated to them.

This is precisely what makes the PGP controversial in genetics circles. Though Church talks about it as the logical successor to the Human Genome Project, other geneticists see it as a risky proposition, not for its privacy policy but for its presumption that the emerging science of genomics already has implications for individual cases. The National Human Genome Research Institute, for example, has cautioned that the burgeoning personal-genomics industry, which includes research-oriented projects like the PGP as well as straight-to-consumer companies like [Navigenics](#) and [23andMe](#) and whole-genome-sequencing shops like [Knome](#), puts the sales pitch ahead of the science. "A lot of people would like to rapidly capitalize on this science," says Gregory Feero, a senior adviser at the NHGRI. "But for an individual venturing into this now, it's a risk to start making any judgments or decisions based on current knowledge. At some point, we'll cross over into a time when that's more sensible."

Church cautions, however, that keeping clinicians and patients in the dark about specific genetic information — essentially pretending the data or the technology behind it don't exist — is a farce. Even worse, it violates the principle of openness that leads to the fastest progress. "The ground is changing right underneath them," he says of the medical establishment. "Right now, there's a wall between clinical research and clinical practice. The science isn't jumping over. The PGP is what clinical practice would be like if the research actually made it to the patient."

In the not-too-distant future, Church says, hospitals and clinics could be outfitted with a genome sequencer much the way they now have x-ray machines or microscopes. "In the old books," Church says, "almost every scientist was sitting there with a microscope on their table. Whether they're a physical scientist or a biological scientist, they've got that microscope there. And that inspires me."

Wired deputy editor Thomas Goetz ([thomas@wired.com](#)) wrote about personal genomics in issue 15.12.

•
•
•

Search Wired

Top Stories

GO

Related Topics:

[Corrections](#) | [Sitemap](#) | [FAQ](#) | [Contact Us](#) | [Wired Staff](#) | [Advertising](#) | [Press Center](#) | [Subscription Services](#) | [Newsletter](#) | [RSS Feeds](#)

Condé Nast Web Sites:

[Webmonkey](#) | [Reddit](#) | [ArsTechnica](#) | [Details](#) | [Golf Digest](#) | [GQ](#) | [New Yorker](#)

<div>Subscribe to a magazin</div>	<div>Condé Nast web sites:</div>	<div>International Sites:</div>
-----------------------------------	----------------------------------	---------------------------------

Wired.com © 2012 Condé Nast. All rights reserved. Use of this Site constitutes acceptance of our [User Agreement](#) (effective 3/21/12) and [Privacy Policy](#) (effective 3/21/12). [Your California Privacy Rights](#).
The material on this site may not be reproduced, distributed, transmitted, cached or otherwise used, except with the prior written permission of Condé Nast.
[Ad Choices](#)