

Hello Ars Technica readers. My name is Rob Reid, and I'm a recovering serial entrepreneur turned podcaster. The best-known company that I started back in the day built the Rhapsody music service, which created the unlimited on-demand streaming model that most people now associate with Spotify. These days, my main job is writing sprawling science fiction novels for the Del Rey imprint at Random House. As an offshoot of my most recent novel - which is called After On - I launched a podcast under the same name. My show dives deep into complex issues in science, tech and society which we should all probably understand a bit better. Each episode's built around an in-depth interview with a world-class expert in the relevant field.

I do 20-30 hours of up-front research and preparation before sitting down with my guests. And I structure my interviews carefully, so that their information density hopefully feels a bit more like TED talk than a meandering long-form interview. Ideally, I try to bring my listeners from a glancing familiarity with the day's subject to a top-percentile understanding of it in the course of the 60 to 90 minutes that most of my episodes run.

We're doing a bit of an experiment here on Ars. Breaking some of my episodes into half-hour segments, roughly, that you can listen to over the course of two to four days while having lunch at your desk, or just avoiding your inbox. We'll run the opening excerpts on Tuesdays for the most part, or Wednesdays occasionally, and the following ones on subsequent days until the episode is complete.

We'll also run each segment with a transcript, for those who prefer reading to listening. As well as a sort of mini-article that will tip you off about the subjects we're covering in the day's installment.

If you like what I do, I hope you'll consider subscribing to my podcast and listening to some of the episodes in archive - all of which were designed to have long shelf lives, and none of which have gone stale yet. You can find my full archive of roughly thirty episodes at after-on.com. Or by simply typing the words After On into the search window of your favorite podcast player. This will also let you hear the entirety of today's episode right away, rather than waiting for the next installment.

Today's episode is the first of three installments of an amazing conversation I was lucky enough to have with George Church. George is one of the most influential people in the worlds of synthetic biology and genomics. He was one of the earliest drivers behind the Human Genome Project. He's one of the most prominent co-inventors of the gene editing technology known as CRISPR. He's been involved in the creation of almost a hundred companies - at 22 of which he's listed as a co-founder. And his Harvard lab is one of the most celebrated fonts of innovation in the world of life science

I begin this episode with something that's quite unusual for my show. Which is four important definitions, that I present before the interview. These should make things a lot more accessible and understandable to non-experts in the field. So if you are pretty deep in synthetic biology, feel free to fast-forward past the next four-ish minutes to the start of the actual interview. For the rest of you, the reason I'm presenting these definitions is that a few years back, the mainstream press began reporting on the gene editing technique I just mentioned called CRISPR -- which is spelled like the English word, only in all caps and without an E.

With this widespread coverage, the term CRISPR has almost become a euphemism for the broader field of synthetic biology. It is indeed a very significant technique, but its overuse as a term can mask certain nuances, which are vitally important to understanding what's happening in genomics today, nuances which are, in fact, vital to understanding CRISPR itself.

So George and I will start our conversation with a higher-level survey of the field, one which will cleanly define CRISPR by placing it into a broader, and also quite fascinating, framework. Specifically, we're going to start by discussing four areas: genetic sequencing, gene editing, DNA synthesis, and DNA assembly.

And now for those four definitions, starting with genetic sequencing.

Sequencing is a fancy and rather cool way of saying "reading." As some of you probably know, your genome is about three billion characters long. It's written in a limited alphabet of just four letters: A, G, C, and T. And if someone sequences your genome, it simply means that they've read it. They've read and recorded that long chain of letters. They haven't modified it in any way. They haven't cloned you. They've just gotten a readout. Kind of like determining your blood type, only a few billion times more complicated. As George and I will discuss, the cost of sequencing genes is plunging with astounding speed, speed which makes Moore's law seem pokey, and the ramifications of this are quite electrifying.

The next area George and I will discuss is gene editing. This is exactly what it sounds like. Editing the genome of a person, a bacterium, or a virus involves changing some of its letters. This can significantly change an organism's function, perhaps causing a small critter to produce something useful, like a medicine or a biofuel, or perhaps someday giving people or animals super powers. CRISPR is a form of editing, but it's not the first one. It's the 10th one. In a lineage that goes back decades. CRISPR's better than most editing techniques at many things, but it's not better than all techniques at all things. And it's definitely not the last form of editing. There's massive headroom for improvement in

genetic editing, and CRISPR will be superseded many times by more powerful approaches in the future.

The third thing we'll discuss is DNA synthesis, specifically the creation of relatively small, customized units called oligos. These are short sequences of DNA, which typically run from a couple dozen letters to a couple hundred letters long. That's obviously tiny in relation to your three-billion-letter genome. It's also tiny in relation to bacterial genomes, which often range in the millions of letters, or viral genomes, which often range in the hundreds of thousands. Oligos are building blocks, which are made to order in specialized labs. Like sequencing, DNA synthesis has gotten radically cheaper in recent decades and continues to get cheaper every year.

The fourth foundational topic is DNA assembly. This is the process of stringing together those oligos into long strands. In theory, there's nothing to stop scientists from linking several million oligos into a strand as long as a human genome, but in practice, errors creep into the process, both as the underlying oligos lengthen from dozens to hundreds of letters and as the number of oligo links in a final assembly rise. As a practical matter, things start getting tricky well before a hundred thousand letters, and the days of a fully-synthesized, error-free, human-length genome are still many years off.

That's the table of contents of the first half of our interview: sequencing, editing, synthesis, and assembly. With those foundations in place, George and I will then talk about the astounding things that this integrated set of rapidly improving and mutually reinforcing fields are enabling. And away we go.

THE INTERVIEW BEGINS

So, George, thank you kindly for hosting me here at your lab on this very blustery day here in Boston. Before we get into genomics, I'd love to talk briefly about your background. You've been here at Harvard almost continuously since the late '70s. Am I right?

George Church: That's right. I was a graduate student from 1977 to '84 and then professor from '86 to present.

Rob Reid: I looked at the list of startup companies whose roots trace either directly or partially to your lab, and it's immense. It looks like it's pushing what? Almost 100?

George Church: Well so, I've been involved in easily 80 companies. I've co-founded on the order of 22, depending on how you count the most recent ones,

which aren't fully founded yet. But usually it was less than one a year for a while. Then it was two, and this year, it's already 11.

Rob Reid: Wow.

George Church: And most of them recently have been post-doctoral fellows leaving the lab, often with other people in tow, starting their own companies. So it used to be that I'd be doing it with some other professor at some other university involved. Now, it's much more homegrown than ever before.

Rob Reid: Wow. And if this new homegrown methodology has spawned 11 companies this year, out of 22 over a span of decades, it certainly sounds like a productive approach. Anyways – as you know, my introduction to this interview will include a pretty robust overview of the topics you and I now plan to cover, which are genetic sequencing, genetic editing, DNA synthesis, and DNA assembly. It'll also include good starting-point definitions of all those things. So that background has already been transmitted, which means you and I can dive right into.

Let's start with sequencing. When I think of sequencing, I immediately and naturally think of the Human Genome Project because of its renown, its size, and its audacity. You were, of course, deeply involved in it, literally going back several years before its beginning. What's your big-picture take on the Human Genome Project now, 15 years after it concluded?

George Church: I personally don't think it was audacious enough. I think it was misdirected, and I'm disappointed. I felt, from the very beginning, in 1984, when we first proposed it, the goal should be to reduce the price, ideally to something that was too cheap to meter, like \$1,000.

Rob Reid: So, to work on the technology and make it much cheaper is a first step?

George Church: Yes. And then, start sequencing genomes. Maybe do a little test along the way, but not aim to spend \$3 billion. Which was our original goal. A number we kind of pulled out of the air – a dollar a base – for one genome. A clinical genome today would be essentially two genomes: your mother's contribution and your father's contribution to you. And so the thing that we did for \$3 billion was one of those, which would not be useful clinically. So we didn't actually produce a useful kind of genome, so that meant that we had to develop a whole new set of technologies afterwards. That I consider the audacious project, which was making a \$1,000 genome, which was diploid, meaning both your mother and your father's contribution.

Rob Reid: This perspective of yours has amazed me ever since you and I first discussed it on the phone a few weeks ago, because I'd always imagined

that the Human Genome Project was a 13-year period in which there was breakthrough after radical breakthrough, and things just got cheaper and faster. But in reality, it sounds like it was quite conservative, in terms of the technologies that were developed and deployed.

- George Church: I considered it crazy conservative. See, cheaper and faster are two very different things. You can always make something faster by just buying more machines.
- Rob Reid: In other words, by spending money faster.
- George Church: Yeah, and that's almost exactly what we did. There were some minor improvements, but to tell you how minor they were, essentially none of the improvements during the Genome Project persisted.
- Rob Reid: So, then, technologies were created during that 13-year period, but they were generally too timid or not transformative enough to last until today?
- George Church: There were a few things before the Genome Project that persisted, like shotgun sequencing. I mean, we still do that. But all the innovations on electrophoresis, which I consider quite incremental – like changing from slab gels to capillaries – those are mostly gone in what's next-gen sequencing. Next-gen sequencing was a complete redo, which was based on microscopy.
- Rob Reid: Just a quick definition for listeners: next-gen sequencing is sort of a blanket term covering a raft of innovations that emerged after the Human Genome Project. In light of all that, what do you think of the \$3 billion investment in the Human Genome Project?
- George Church: Basically, that was thrown away, because neither was the genome of a clinical grade, nor were any of the technologies preserved.
- Rob Reid: To put the power of post-Genome Project technology – in other words, next-gen technology – into perspective, the project itself spent \$3 billion over 13 years sequencing a highly-imperfect genome, and it ended in 2003. How does that compare to the cost and time requirements of sequencing a much more robust human genome today?
- George Church: Probably the lowest price right now is \$600. The timing depends on how you count the DNA prep and various other things, but it's short. It's in days now rather than in years.
- Rob Reid: That's well over a three million-X improvement in price over the 15 years since the Human Genome Project ended. A time period which is

almost equal to the time span of the project itself, interestingly. So what do you think would have happened if the focus had been on the technology first? Do you think we would have gotten to the end point of the Genome Project much more quickly?

George Church: Oh, yeah.

Rob Reid: Much more quickly?

George Church: Yeah, and we would have had \$3 billion to generate genomes. We still haven't spent \$3 billion on clinical-grade genomes.

Rob Reid: In the time since the end of the Genome Project?

George Church: In the time since then. Because we kind of burnt it up on the first one.

Rob Reid: So, why did it play out that way? It sounds like you, for one, clearly wanted to develop the technology first.

George Church: I didn't have much of a vote. I mean, I was the youngest person in the room in 1984. And I was probably the youngest person with a grant when the grants started getting handed out in 1987. I think the reason wasn't evil or anything. It was just that there were very few engineers at the time in biology, and the few engineers had no clue about molecular biology. So they might be engineers working on artificial joints or something – or maybe engineers that could build a better patch clamp for neurobiology. But very, very few in molecular biology.

Rob Reid: And so when the Genome Project ended, there was – what you recently described to me in a prior conversation – a “technological overhang.” A lot of developments, which had been partially made under the radar, in skunkworks, burst into the light and started getting fully fleshed out. And that's why we've had this crazy acceleration since the end of the Genome Project. Where does it go from here? Do you see another three to five million- X improvement over the next 15 years, or are we going to run into the limits of physics?

George Church: Well, ultimately there will be some limits of physics, and you can get some idea of that from how quickly and cheaply you can transfer data from one form to another. Like transferring it from Blu-ray to a RAM in your computer.

Rob Reid: In other words, we should one day be able to transfer a genome's worth of DNA into digital storage as quickly as we can transfer from one digital medium to another. What does that say about the eventual cost of sequencing a genome?

George Church: Transferring a genome equivalent, which is less than a Blu-ray disc, into your computer is a penny. It's probably the electricity plus the amortization of your laptop. That's where we're headed, is my guess.

Rob Reid: So, we're really in a headlong rush to the dollar genome, then to the penny genome, and maybe even to the sub-penny genome.

George Church: And long before we get there, consumers will be seeing \$0. And maybe even being paid to get their genomes sequenced.

Rob Reid: Because that data will be valuable enough to other entities in the system to merit that?

George Church: It's terribly valuable already. We've already gone past the point where all of us should be paid to have our genomes sequenced, because the system could make on the order of a million dollars every time we save one child from a serious Mendelian disease. That million dollars should be then spread out to all people who participated, including all the people who didn't get any bad news.

Rob Reid: In other words – if detecting a disease early can nip it in the bud and save the healthcare system a million bucks, if just one person in 1,000 has a detectable disease, even with today's cost structure, we're already ahead of the game if we sequence everybody. Then certain diseases can be avoided entirely if two people who both carry a dangerous recessive trait figure that out before they create a child.

George Church: If you get the carrier status, you can avoid other carriers of the same thing. And to some extent, at that point, you and all the people that didn't get carrier status should be equally compensated.

Rob Reid: All this takes us well beyond the definite article in **THE** Human Genome Project when it was really **A** human genome was being sequenced.

George Church: Not even.

Rob Reid: Not even. It was half of one, right. So let's talk about what happens when we go from a human genome to hundreds of thousands, and eventually millions, of complete genomes in a database. There are now several projects underway to create giant databases. One of which you initiated, the Personal Genome Project. What do you imagine we'll be able to achieve with the vastness and the depth and the richness of the data that we'll have, let's say, in 5, 10 years with millions of genomes sequenced. And also, ideally, matched to phenotype and health records?

George Church: Well, the critical thing is that coupling with phenotypic information – the health records, and big environmental component of the microorganisms and viruses you're exposed to, and how your immune system responds to them. If you have all that, almost everything we die of is genetic. Now, some of it is both genetic and generic. We're all going to die of age-related diseases, at least the lucky set of us that live in industrialized nations and escape infectious diseases. But we will get insights into all of human biology. And many of the solutions won't depend on us necessarily finding the cause and reversing that cause. Some of them will have completely synthetic approaches. In fact, most of the pharmaceuticals, really, are not reversing some natural process, literally. It's coming up with something kind of out of left field that's used at the pharmacological level, not a physiological level. Synthetic biology is a better paradigm than natural population genetics or natural human genetics.

Rob Reid: But natural population genetics will give us major insights because we'll start being able to do much, much bigger correlation studies. Of course, we've already correlated thousands of genes with thousands of conditions. But those correlations aren't perfectly understood because we don't yet have millions and millions of genomes. And we certainly don't have the match to phenotypes and health records. But when we do start getting into the millions of genomes sequenced, is there a point at which our knowledge becomes so statistically significant that there's very little value to adding still more genomes to the database? I mean, is there a threshold beyond which we kind of have enough genomes to understand human genomic? Or if it's possible to sequence the next and the next and the next person, it'll always be scientifically beneficial to do that?

George Church: We have 7.5 billion people on the planet – I don't think that's enough. I think there's a good argument for having more babies, so we get more insight into human biology. Now, that's a little facetious – but the point is, it's not just about correlation. It's about cause and effect, and we can learn a lot from one person. In a certain sense, if you have one person you can test cause and effect. You could follow up on anything that looks bizarre about that person, about their genetics. And we're getting better and better at determining cause and effect, and that may, in the end, be more powerful than correlation from billions of people.

The thing that's interesting about having 10 billion people or 20 billion people as your resource is that you can find really rare things which are N of 1. They really interesting N of 1 cases where they've had an unusual environment, or they've had some weird combination of alleles, and that makes them unique. That's what I think is the advantage of having more people, not necessarily moving the number of times I've observed a particular allele from 1,000 to 10,000. That points to diminishing returns.

Rob Reid: If you find an N of 1, a unique, one-of-a-kind case we might discover upon sequencing, let's say, the seven billionth person, the scientific value wouldn't be in discovering yet another rare disease, right? I mean – clearly you'd want to cure that person if you did find a disease in them. But from a public health standpoint, that disease would be far too rare to move the needle, right?

George Church: I'm less intrigued by the N of 1 being a rare disease we're going to cure. I'm more intrigued by the N of 1 being a rare protective allele that's going to save us all. For example, you can find incredibly rare people who are resistant to HIV. Some of them are resistant because they don't have receptors on their T cells, and some are resistant because they've got a fluky antibody. My lab participated in some of these studies that studied these exceptional people that escape HIV. Those antibodies could be powerful protection for the rest of the planet. Or you get people that live to 122 years old. Maybe they have something that all of us could share. Maybe there's a dozen of them that each have a different thing, and each of those N of 1 gives us one more tool in our effort to reverse aging. So that we have, essentially, the body and mind of a 22-year-old, but the experience of a 130-year-old.

Rob Reid: That's why the hypothetical 20 billionth person could be the source of a great discovery. Now, when I first started thinking about personal genomes, maybe 10 years ago when 23andMe started getting a lot of press. And I started realizing, "Wow. Some day we're going to have a \$1,000 genome, and then a \$100 genome." Back then, I assumed there'd be some kind of end point beyond which cheaper and cheaper sequencing wouldn't be very useful, because what do you do once we're all sequenced? But, of course, it turns out that there's lots of useful things that go far beyond personal genomes, which can only be enabled by ever-cheaper technology. I'd love to discuss something you've speculated about maybe in the 5 to 10-year future, which is genetic monitoring – a system that could let you determine if there are unfriendly pathogens in a room as you enter it, or even before you enter it. What kind of device and horsepower would we need to pull that off?

George Church: My gut is that it's going to be some electronic sequencing. So, either nanopore, or some other single-molecule electronic, which has very low reagent costs, very small fabrication costs, possibly portable, possibly disposable. And you could scatter them around the room, you could scatter them around your body – hey could be that inexpensive. And then you've got a social network that's exchanging information about where all the pathogens are in various rooms, in various parts of the world. And it would really revolutionize epidemiology because it'd change it from being something where you do a 10-year study to something where you do a 10-second study. And you'd know whether you're sending your kid to daycare with H3N2 – some deadly flu virus –

or norovirus, or Ebola, rather than some minor rhinovirus or coronavirus, that's just going to cause a little sniffles in everybody in the class.

Rob Reid: This would be enabled by extremely fast, extremely small, extremely cheap sequencers which would basically inhale pathogens out of the air and identify them. How many years off would that level of robustness be? Are we talking decades, or is this in the more intermediate future?

George Church: I think almost nothing we're talking is decades. We're already at the point with nanopore sequencing where you can put in samples and start getting answers out in 25 minutes. The DNA preps are really easy, so it's increasingly integrated with the device. The step from that, to sampling something from the air – which tends to be pretty clean – you would get pollen, you'd get some viruses, you'd get bacteria. All that would go into a simple deproteinizing and into the nanopore. I think we're talking about years, and the software's not hard either.

Rob Reid: So, single digit years?

George Church: Yeah.

Rob Reid: Interesting. If you can already do this kind of test in 25 minutes, the path to 10 seconds seems daunting, but that's only what? Two orders of magnitude in a field that routinely advances five or six orders of magnitude in a decade or two?

George Church: Yeah. Speed you can gain by parallelization. The process of stripping proteins from a virus is pretty fast. It really can be instantaneous. And then it threads into the nanopores instantaneously. That is very close.

Rob Reid: And device size could be something that just sits on your smartphone or hangs around your neck?

George Church: The device size already that thousands of people are using is 70 grams, the size of your cellphone.

Rob Reid: Obviously, that's going to shrink. On a kind of related note, you know Danny Hillis, right?

George Church: Yeah!

Rob Reid: Are you familiar with this app he created called Dark Sky?

George Church: Hmm?

Rob Reid: It's a cool weather app which leverages the fact that most smartphones actually have barometers built into them. So the millions of phones that

are running Dark Sky constantly report back barometric readings, and those readings are tagged with GPS coordinates, which lets the system give you a reasonably accurate minute-by-minute forecast for your tiny micro-region. Like maybe a 100-yard circle, I don't know. In other words, it also leverages detectors that millions of people in the network are carrying.

George Church: Well, there's a similar thing for airplane navigation. And, of course, for car navigation – Waze.

Rob Reid: Exactly. Your detector network would be like Waze or a Dark Sky for disease. And it would get more and more powerful and accurate as more and more people adopt it. That's amazing. What do you think the 2.0 version of this might do, as if that's not already amazing enough?

George Church: Well, I think one thing is preventing us from spreading known pathogens. The other is discovering the next pathogen before it's an epidemic. You can notice that it's going through birds or pigs or something like that. They're detecting your immune response to it. Steve Elledge and I worked on methods for checking your autoantigens, and that could be something that's close to real time. And so you could start connecting, "What do I get exposed to that causes me to become autoreactive?" Because I think a huge number of diseases that we have that are not well explained right now are actually our interaction with our environment and our immune system. Probably a lot of things that cause chronic fatigue or arthritis, gastrointestinal problems, diabetes, and so forth are complications where part of it is genetic, but a huge part of it is immune response to the environment. I think that we could get much better at that.

Rob Reid: By doing some sort of monitoring inside the body?

George Church: Well, actually, the way that Steve and I did it is you take a lance, just like diabetes tests, and you need much less than a drop of blood. You can do this whole assay for every possible virus, every possible human antigen. But yeah, internally would be even better. In fact, it'd be really great to embed a lot of sensors because it's all getting to sub-millimeter circuitry, so there's no reason why you couldn't make something that's very noninvasive, but is monitoring 24/7.

Rob Reid: Now that we've talked about sequencing, let's get to part two of our discussion: editing. Editing will continue to matter hugely until the somewhat distant day when we can make arbitrarily long error-free DNA strands, because editing lets you start with an approximation of the end point you're targeting, assuming that approximation's already largely built into the genome of a virus or bacteria or an animal or whatever. Then you just have to make relatively small changes to reach your goal. CRISPR's gotten a huge amount of press and adulation as an

editing tool, but it certainly isn't the first genetic editing tool. Could we talk a little bit about the heritage that preceded CRISPR, and why CRISPR is a significant breakthrough – but also why it's not the last breakthrough, and why there's a lot of headroom for improvement from here?

George Church: CRISPR looks to me to be a momentary pause rather than a gigantic breakthrough. And I can say that because it's partly my baby, a shared experience with many others. The earliest efforts in mammalian cells, at least, were groups like Mario Capecchi and Oliver Smithies, who would introduce double-stranded break linears – into say, mouse cells – and they would, at some low frequency like one in 1,000, one in a million, get incorporated in at the right place. Then you'd have to screen through those, and then you could construct a mouse that was exactly what you wanted.

That 1980's technology, which got a Nobel Prize, was better than CRISPR in certain ways, in that a mouse was exactly what you wanted. What most people call CRISPR editing is knocking things out. It's not precise editing. It's kind of bashing. Along with and after Smithies and Capecchi, then you had a series of enzymes that would help you soften up the genome a little bit, so moving it from one in 1,000, one in a million to more like one in 100. Those were meganucleases, which Bernard Dijon, one of my advisors, discovered in the '70s. Then zinc fingers and zinc finger nucleases. And then TALs and TAL nucleases. Those all would cut, but there was a race between fixing the cut with what you wanted – a perfect edit – and just making a mess. Then there was another set that was much better at doing what you want without making a mess. In other words, instead of making a double-strand break, they would swap in the DNA. There would be more dance-like cooperation between the donor DNA and recipient.

Rob Reid: So, it was more of a pasting operation than a cut-and-insert operation?

George Church: Exactly.

END INTERVIEW ELEMENT OF SEGMENT ONE

Hello again, Ars readers. On that cliff-hanger we conclude this, the first excerpt of three from my interview with George Church. Which, if you're curious, originally ran on my podcast on April 3rd. If you can't wait to hear the rest of it. Or, if you'd like to browse my other 30-ish episodes, you can just head on over to my site, at after-dash-on-dot-com. Or, type the words After On into your favorite podcast player. You'll find a lot more stuff about genomics and synthetic biology. Multiple episodes connected to neuroscience and consciousness. Conversations about robotics, privacy and government hacking, cryptocurrency, astrophysics, drones, and a whole

lot more. Or, you can join me here tomorrow on Ars, when we'll continue with Part Two of this interview.