AI GR Podcast 11 10.19.23 George Church

[00:00:00] And so we put in two pages on ethical, legal, social implications, and how we were going to do the IRB. And in the process, this is the first time I've actually done an IRB. I'd collaborated with people who had done IRBs before. This is the first time I did it. And I was just appalled at some of the things that they were forcing you to do.

[00:00:20] They, they said like, you have to promise that you'll keep it private. And I said, I don't know how to keep my credit card private, much less somebody's medical records. And in fact, there had been numerous, uh, hacking events and people losing disk drives at various hospitals in the Boston area that will remain nameless.

[00:00:41] And so I said, I think the better assurance would be that we get people who are okay with it being public. That was one thing that I was horrified by. The other one was they said you couldn't return information to the patients, even if it could save their life. And I felt that was crazy too. And I [00:01:00] said, well, but a consequence of their data being public is it also be available to them.

[00:01:04] And so they can. You know, if we give them the right software, they can draw their own conclusions and with proper consulting with their physicians and specialists, they will make the right decision. And so, we framed it as an ethics study as well as a technology study, and that became the Personal Genome Project.

[00:01:26] Welcome to another episode of *NEJM AI Grand Rounds*. I'm your host, Raj Munrai, and I'm here with my co-host, Andy Beam. Today, we're delighted to bring you our conversation with George Church of Harvard. George is a professor of genetics at Harvard Medical School and professor of health sciences and technology at Harvard and MIT.

[00:01:44] He has many other titles and honors. Andy, I was really struck by both the depth and the breadth of cutting-edge science that George oversees in his laboratory, and also by how effectively he must context switch. He's very well known for his pioneering work in genomics and synthetic biology, [00:02:00] his involvement with the Human Genome Project and the Personal

Genome Project, his work on gene editing, and even revival of the woolly mammoth, just to name a few areas.

[00:02:09] He also shared aspects about his career journey and how he approaches both academic and entrepreneurial efforts, which were enlightening. You know, Raj, George has to be one of the most fascinating individuals I've ever met. I think when most of us decided that we wanted to be scientists, we actually probably wanted to be someone like George.

[00:02:25] He's always working on the most intriguing things, and I'm constantly amazed by the range of truly sci-fi projects that he has going on. I really love that he sees his mission as making the future happen faster. You know, and that way he really reminds me of a scientist from a movie almost. He's got this big beard, this big personality, and he does crazy things like try to bring back the woolly mammoth.

[00:02:45] I always enjoy hearing about what George is up to and this conversation was no exception. The *NEJM AI Grand Rounds* podcast is sponsored by Microsoft and Viz AI. We thank them for their support. And with that, we bring you our [00:03:00] conversation with George Church. Welcome to *AI Grand Rounds*, George. We're happy to have you here.

[00:03:06] Ah, it's great to be here. So, George, a question we ask all of our guests. Tell us about the training procedure for your own neural network. How did you get interested in science, and in AI in particular? What data and experiences led you to where you are today? I've heard there is a frameable Duke letter as part of this story.

[00:03:26] Yeah, so I got interested in science, both natural and unnatural. My third father was a physician, and so I was fascinated by the technology in his... He would do house calls back then. Bit of history there. And then I went to the World's Fair and saw... Computers and robots or animatronics. And I decided that was definitely cool.

[00:03:49] And I came back to Florida where they didn't have any science classes to be taken, that I could find. So, I just started building my own computers because I was desperate, not very good ones. [00:04:00] And so it wasn't until ninth grade that I actually got access to a GE 635 at Dartmouth. Been programming ever since, basically, that was actually a time sharing interactive one.

[00:04:10] This is in 1968. I'm in ninth grade. Then when I went to college at Duke, it was kind of a step backwards, forwards in terms of science. I was doing crystallography, but backwards in terms of, uh, everything was punch cards into IBM 360. And then I got two major degrees in two years and then proceeded to flunk out of graduate school.

[00:04:31] So they gave me a nice letter, hoping that I would do well in some other field, but I stuck to this field and, my lab has been mostly, mixture of technology development and computational biology pretty much since the beginning in 1986 at Harvard Medical School. And naturally as the names kept changing as to whether it's neural nets or deep learning or machine learning, but we've been part of that revolution, at least as applied to things like molecular biology,

[00:04:59] [00:05:00] protein design in particular. Thanks. That's super interesting. And I actually didn't know that you had such an early interest in computer science. And I can see that writing code, whether it's in bits or in DNA has sort of been with you since the beginning that you're still programming the software. The substrate has just changed a little bit over time.

[00:05:17] Yeah, yeah. My first computers were an analog computer and then a digital mechanical computer. So, it was only the third one that was, was a real von Neumann, digital interactive computer. We're going to hop into your research in just a little bit, but there's an icebreaker question that I think I'm obligated by law to ask you, and it's about bringing back the woolly mammoth.

[00:05:40] And, so, you've been on record on, the record as saying that we should bring back the woolly mammoth. And I'd love to hear, the motivations for doing that and sort of how far are we away from being able to do that now? Well, this is not going to fit into 140 or 280 characters, but I'll try my best.

[00:05:58] So it's really not about [00:06:00] bringing back an extinct species, we're mainly interested in endangered species and restoring ecosystems. The ecosystem that seems like it needs the most help is the Arctic, where there's a lot of carbon sequestered because of the cycles of freezing and then summer photosynthesis results in sometimes 500-meters thick of topsoil compared to say 1-meter in a lot of the rainforests.

[00:06:26] And that's all melting and it's melting a lot of it in the form of methane, which is 80 times worse than carbon dioxide. And so, we're hoping that making cold-resistant elephants, which would help save them from their

endangered species categorization and would also help restore that when we're not saying this is the only answer or even

[00:06:48] this is a good answer. It's just our best shot at increasing sequestration and decreasing the loss in the form of methane. And we're doing that by genome [00:07:00] engineering, which we've already done in pigs for making transplants, xenotransplants, which are making their way through now 600 successful days in non-human primates and some.

[00:07:14] And now going into human clinical trials, we can do dozens, uh, 42 edits in pigs genome germline. We hope to be able to do the same thing in elephants and make them cold-resistant, maybe also resistant to herpes viruses that are killing them. If you have to do 42 edits for pigs, what is the difference in scale to do something to like turn the elephant into a cold-resistant woolly mammoth?

[00:07:42] A cold resistant elephant, that has genes that are resurrected from woolly mammoth. We've already resurrected two genes that seem to have the right properties when resurrected. We don't know the exact number. We're doing a computational analysis of dozens of [00:08:00] elephant and extinct relatives and focusing on genes that go to fixation.

[00:08:06] That means that they're homozygous, that both their maternal and paternal alleles are the same, and both of those are different from the existing elephants. And so basically, you've got this branching in the phylogenetic tree where it's gone to fixation where all of the mammoth genes are different from the existing ones.

[00:08:30] And we can do that for multiple mammoths and multiple existing genomes, both African and Asian elephants. And then we're focusing on those, both the coding regions and the non-coding regions. And if we did all of them, it would be in the hundreds of thousands. If we just do the ones that we think are involved in cold, a priori, well then those could be in the dozens.

[00:08:54] So, it's somewhere in between. And we have technologies, editing technologies, I think that are up to the [00:09:00] high end of the spectrum. Our record so far for repetitive elements is editing 24,000 edits simultaneously in one cell. That happened to be a human pluripotent stem cell, but it could have been any mammalian genome as far as we know.

[00:09:16] Got it, thanks. So, George, I want to switch gears to your work in genomics. You're a pioneer in genome sequencing, genetics, synthetic biology,

many other areas. And your work has been influential on many scientists, myself included, but I actually want to ask you about a different perspective, which is your perspective as a patient.

[00:09:37] You famously made your own genome sequence, your lab values and medical records publicly available for anyone to download on the Internet. I want to ask: How did you come to that decision and what you might have learned since you've made that information available? Yeah, it's an interesting question.

[00:09:55] So we, uh, applied for NIH Center of Excellence [00:10:00] in Genomic Science and got it. And in the process, we were proposing to develop a new way of doing DNA sequencing. Now, this is in 2003 and we were partway there anyway. So that's why we proposed it. It's now called NextGen sequencing, but back then it had other names.

[00:10:18] And we proposed that in the process of the five-year Center of Excellence grant, we would get as far as doing a 1.7 million base pair bacterial genomes. This is a very tiny bacterial genome, not the smallest, but very tiny. It was Helicobacter pylori, the causative agent of isthemic cancers and ulcers. And we had already contributed to the first time that was sequenced back in 1994-ish.

[00:10:46] So, in 2003, we proposed to do that again after five years of work. As it turned out, eight months into it, we had finished something three times bigger than that, and by the end of the project, we had finished [00:11:00] five human genomes at six billion base pairs each, so diploid, high-quality genomes. Now that it was not completely unanticipated, we thought we might get the human genomes, even though we were only promising to do a really tiny bacterial gene, a thousand times smaller, we thought we might get there.

[00:11:15] And so we put in two pages on ethical, legal, social implications and how we were going to do the IRB. And in the process, this is the first time I've actually done an IRB. I've collaborated with people who've done IRBs before but this is the first time I did it. And I was just appalled at some of the things that they were forcing you to do.

[00:11:33] They said like, you have to promise that you'll keep it private. And I said, I don't know how to keep my credit card private, much less somebody's medical records. And in fact, there had been numerous, uh, hacking events and people losing disk drives that various hospitals in the Boston area that will remain nameless.

[00:11:53] And so I said, I think the better assurance would be that we get people who are okay with it being public. That was [00:12:00] one thing that I was horrified by. The other one was they said you couldn't return information to the patients. Even if it could save their life. And I felt that was crazy too. Uh, and I said, well, but a consequence of their data being public is it also be available to them and so they can, you know, if we give them the right software, they can draw their own conclusions and with proper consulting with their physicians and specialists, they, they will make the right decision.

[00:12:27] And so, we framed it as an ethics study, as well as a technology study. And that became the Personal Genome Project. That's fascinating, that history. I'm curious, so you also, as part of this, you made your own genome public. Oh yeah. And I think, to this day, you can go and download a copy of your genome as well as your medical records.

[00:12:50] And just connecting the dots, you were one of the participants who was willing to have their genome released, right? You were a participant in your own. That's correct. The IRB [00:13:00] felt that, I mean, it was an unusual proposal to the IRB because I was breaking at least two of their rules, a few others that we don't need to go into, but, but, you know, transparently and politely, and it took us about a year to negotiate it, but that's not that long compared to other IRB approvals, even non-controversial ones. But part of that is they said, well, you know, you should be willing to eat your own dog food.

[00:13:23] Uh, you know, we want you to be the sole participant. And I said, well, how about 10 participants? Let's compromise. And I'll be the first of the 10 and then we'll expand it. If nobody gets hurt, we'll expand it from there. And John Hovka was actually number two. Uh, Esther Dyson was number three. All their names were known.

[00:13:42] That was approved by the IRB as well. And John was at the time just brand-new CIO at Beth Israel Deaconess and also had crafted the, I think it was Norway's government's medical informatics policies, and so, he was really perfect second. In fact, there's a funny story about [00:14:00] my posting. So, I was a patient at Beth Israel Deaconess, as it turned out, and before we had the IRB approval, I had just posted my own medical records as a test, and some patient had been looking, browsing the Internet, and had run across my medical records, and she freaked out.

[00:14:16] She thought, this is standard policy at Deaconess, was to release patient medical records onto the Internet, and her complaint worked all the way up to the president. And then back down to John Holomka, who said, oh, I

know what this is, this is George Church. And he contacted me, and I said, oh yeah, yeah. Oh, I didn't even think that some patient would find this.

[00:14:36] I said, put a big disclaimer on it. I said, this is not Beth Israel policy. This is just an experiment. And then they thought that was okay. And we went forward. And so, have you learned anything about your own genome since it's been publicly available or your own medical records since you've made these widely available?

[00:14:57] Oh, yeah, sure. There have been various [00:15:00] advantages. There was one case where I was in giving a lecture. in Seattle and somebody in the front row said while we were waiting for everybody to settle in, you know, there was kind of a pause and he said, you know, you should get your blood work checked out because according to your public medical records, you're on statins and there's no evidence that you've checked on your cholesterol or on the possible negative consequences for muscle damage.

[00:15:30] And I said, yeah, you're right. I haven't been checked. And so, I went off and I checked and actually there was no lowering of cholesterol and there wasn't muscle damage, so, or potentially the biomarkers for that. And so, we changed the statin and the dose and monitored the, that probably, you know, gave me an extra 10 years of life for not using drugs right.

[00:15:55] And then, you know, I learned that I had, risk factors, for Serpina 1, for Alpha 1 [00:16:00] antitrypsin, which put me at risk for a whole variety of, respiratory. Problems. And so, I've been cautious about living in, downtown Beijing and Los Angeles and things like that, or, or hanging around with COVID too much.

[00:16:16] Great. So, we're going to switch gears just a little bit again and talk about some of the work that your lab has been doing in artificial intelligence and deep learning for biotech applications. So, I've, you know, I tried to pick one paper to focus in on, but you've done so much work on things like protein engineering, designing AAVs for gene therapy, improving the efficiency of CRISPR.

[00:16:38] So, I was hoping you could pick one from the list of the many papers that you've written in this area and help us understand what machine learning is bringing to the table. So, I think there's often a lot of Confusion about when and how and why to use machine learning, but I was hoping you could really help us zero in on the types of new questions that machine learning has [00:17:00] let your lab ask and answer.

[00:17:02] Yeah. So, the one that I would pick, or the two, the first one was. The one that I would pick is, is a *Nature Biotech* paper we published in year 2000, which is Bryant et al. Eric Kelsick, uh, was a postdoc in my lab, who was a senior author on that paper. And it was titled "Massively parallel deep diversification of AAV casted proteins by machine learning."

[00:17:27] And we have been working on machine learning for protein design for quite a while at that point, but that was different in that it illustrated your question. Your point was when is the time to do machine learning? When is it not? Typically, there's one prerequisite for machine learning is you have a lot of data.

[00:17:45] We had just published some paper that we refer to as low-end machine learning, but it isn't entirely low end, and there's some additional background information about proteins in general, but low end for your particular experiment. But, [00:18:00] but anyway, that's the prerequisite. You need to have a lot of data, typically.

[00:18:03] Then there's the question of when is it better? So, in this particular case, I think we showed that it's a lot better because we did a comparison of naive or random, semi random models for mutagenizing a stretch, a key stretch of 28 amino acids that's important to the AAV capsid for gene therapy delivery.

[00:18:26] And we asked, you know, how many amino acids can we change simultaneously? And the reason you might want to change a lot of them simultaneously is that's where the immune system interacts with it. And if you want to use the gene therapy more than once and you want it to be or even once, you want to make sure the immune system doesn't attack your precious therapeutic.

[00:18:45] Also, if you want to target a new tissue, you might have to radically change the surface of the virus so it will target tissue A and not B. So, those are objectives and we, in the naive model, it was [00:19:00] very hard to get more than four changes out of 28. And this was very consistent with all the work we've done and many other proteins before.

[00:19:07] Four out of 20 was actually a pretty good day in one round of diversification and selection. But using logistic regression or various concurrent neural nets or various neural nets, uh, we could get up to, we could get about 90 percent at 25 out of 28. About, just doing this from memory, 70, 80 percent at 26 at, and even 20 percent of this big library, focused on 28 out of 28.

[00:19:36] So, it wasn't like one solution, it was lots of solutions. Well, at four out of 28 with the naive model, you are getting close to 0%. So that to me is a dramatic validation that this is an improvement over the naive models. Yeah, so this all started with an undergraduate in my lab, Harvard undergrad, Ethan Alley, and Serge Biswas, and others,[00:20:00] Pierce Ogden, who had developed UniRep, which was a language model.

[00:20:05] So, there's a kind of two broad categories in protein structure and design. One is focusing on the structure. And that got a lot of attention with AlphaFold from DeepMind Google, but it just predicts the three-dimensional structure of the protein. Just, I should say that's a 50-year holy grail, but from a protein design standpoint, knowing the three-dimensional structure

[00:20:27] is just a starting point and not necessarily the right direction. So, for example, I'm just going to end on this example, just contrasting the language model, which is essentially linear with the 3D structure, is if you, if you have a serine protease, which is named because serine at the active site is critical, and you change that serine to an alanine, so now it's completely dead, but it has exactly the same three-dimensional structure.

[00:20:53] Now there's the version of this alpha fold, which is alpha fold multimer, where you can ask whether two [00:21:00] proteins can stick together, and that might save you in certain circumstances. But basically, point is you can kill a protein without changing its three-dimensional structure, but not in the language model. It's too, it's sophisticated stuff.

[00:21:12] Right, right. So, that I think dovetails nicely into a follow-up question that I'd like to ask is: you've seen many revolutions in biology, I think it's fair to say genomic sequencing, synthetic biology, could you help us understand how AI is going to make an impact in biology over the next 10 years? And the sort of flip side of that is, are there areas of biology, are there corners of biology that you think are immune from sort of the AI incursion that we're seeing today?

[00:21:44] Yeah, so, uh, I kind of have a policy not to say anything's impossible because it's one way you can show how foolish you are. I mean, I'm not questioning whether I'm foolish. It's hard to prove a negative. Yeah, and, and it's easy to get [00:22:00] embarrassed, two years later when somebody shows that it does work.

[00:22:02] So, and in fact, almost everything that I've worked on, every project that worked on in my 68 years has been, 50 years as a scientist. One person or

other has said it's impossible. Sometimes a couple of years after I published peer reviewed paper on it, they still say it's impossible. But anyway, so I, I can't think of anything where, you know, the laws of physics or computing would prevent it from making a

[00:22:31] contribution. May I qualify? Is there an area of biology you think has low probability of being impacted or is just too hard for AI over the next 10 years? So, I've relaxed the condition a little bit. Yeah, I still think it's, I, I think they're all sufficiently high probability nonzero, uh, that they, that they're worth considering.

[00:22:51] I think things where it's really hard to get data, but on the other hand, it's not clear any kind of intelligence, artificial or otherwise, [00:23:00] is going to solve some of those problems. So yeah, I think there's a good room for it. It doesn't mean that human beings aren't going to play a gigantic role in nudging it, programming it,

[00:23:13] interpreting it, interfacing with people who don't need to know all the, the gory, uh, computational details. So yeah, I think it's going to affect everything really. And I, and I think the same thing is true for synthetic biology. Synthetic biology is going to affect everything, including making computers that might be better than current computers.

[00:23:37] They'll probably be hybrid computers of various sorts. In a certain sense, machine learning is based on, is inspired at least by, um, natural neuronal computers, and I think hybrids will have a good shot at it. If it was anyone else, I would say that was a hedge, but when George Church says never say never, I believe it 100%.[00:24:00]

[00:24:01] Thank you. So, George, I want to ask you about... Is there a field where you think it won't work? Or has it worked? I've just heard a luminary of the field, uh, advise me against, uh... I should have asked you first.

[00:24:17] So George, I want to ask you about commercialization. So, you've been involved in many companies. I think a few that are relevant to protein engineering and AI include Manifold Bio, Nabla, Dyno Therapeutics. Could you tell us about what these companies do and maybe how their missions relate to one another?

[00:24:37] Right, so the three that you mentioned, that Nabla, Diyo, Manifold, and two more, Patch and Shape Therapeutics, all are using machine learning for protein or nucleic acid design. They're wildly different. So Nabla is focusing on

antibodies, which are one of the key therapeutic categories and have diagnostic uses as well.[00:25:00]

[00:25:00] Dyno is on delivery. We were just talking about AAV. Manifold is very interesting and they're developing ways to make protein binding pairs, but that's not the end game. Then those can be used for testing multiple protein therapeutics simultaneously in one, let's say, expensive test animal for preclinical trials, let's say a non-human primate.

[00:25:27] So, you can do thousands of simultaneous measurements of pharmacokinetics and dynamics and tissue targeting and so forth, all at once with protein therapeutics, which don't normally have a nucleic acid barcode that you can follow. So, this is a protein barcode. And then Shape is working on RNA therapies, tRNAs, and ADAR, and also delivery.

[00:25:52] And then Patch is on cis regulatory elements, DNA and RNA. And there could probably be about 10 more of them that would not overlap one another. [00:26:00] Uh, it's a, it's a very important, subset of things that you can do with machine learning. So, you know, as you think about those companies that are making use of machine learning and artificial intelligence and the others that you've been involved in, I'm curious, have you identified maybe a set of, let's say one or two key questions that allow you to decide whether an idea is right as a commercial entity or whether it maybe belongs in academia for more development before it moves into, into a commercial entity?

[00:26:34] Yeah, this is a really tough call that, that every postdoc that wants to start a company and the PI that maybe wants to start it with them or wants to jump from academia into industry. And I say, it's not so bad to be number two, to be the second one in the field, even if you thought of it first and it kind of

[00:26:56] irritates you that somebody jumped in there with your idea. You [00:27:00] just want to make sure you've got enough intellectual property that you have freedom to operate. And it doesn't really matter who gets funded first necessarily. And very often the second one does a better job of it. Either from a business side or a science, engineering side.

[00:27:16] So that's part of the decision. The part where you know that it's ready is either you're getting a lot of feedback from your peers. They like it. They want it. One case, you know, when we were developing DNA synthesis, the very high throughput DNA synthesis, like 10, 000 times previous throughput. And suddenly we had a lot of friends that wanted to buddy in and collaborate on making big

[00:27:42] DNA constructs cheaply. And we said, oh, this is going to be a tremendous academic distraction. We really have to spend this out just for our own sanity. So, we did that same thing with, with CRISPR. That was clearly going to be so popular. I mean, we, we announced it in January 2013, and by March [00:28:00] there were like 10,000

[00:28:01] users, we distribute it through Addgene, which is nearly free, and that would have been a real hassle to have done that. Licking stamps in our office. So, those are the kind of things that tell you that you're ready. You don't have to rush. You can, the more mature it is, the slower you will get diluted out.

[00:28:22] Where, if you go there too early, you may think it's magic money, it's easier than a grant, but pretty soon it'll be taken away from you and, and it may not go in the direction you want it to go once you lose control over it. I, I like the emphasis on it's not so bad to be second. Uh, there's a saying that I'm reminded of that the early bird gets the worm, but the second mouse gets the cheese.

[00:28:44] And that sometimes being second is strategically good. Oh, right. That's good. Second mouse gets the cheese. Second worm gets to live. Yeah. Right. Yeah. Yeah. Right. So, so George, just as a follow on to that, having [00:29:00] overseen many companies from genomics and from sequencing and genomic technologies, and now increasingly with AI, is there a difference you see in the decision making of when to start a new company in genomics in the context of your involvement in companies there versus now with the more AI focused companies or the, the principles more or less the same?

[00:29:22] Well, there's some very significant differences. Well, first of all, when I started back in the 80s, I was basically just the way I was deciding was if an investor or a colleague came forward and said, hey, we'd like to collaborate with you on a company. And I'd say, okay. It was fairly reactive, but recently it's been mostly postdoc driven.

[00:29:43] And during that time, we transitioned from mostly analytic to mostly synthetic. And the problem with a lot of the DNA sequencing scenarios was you had to convince people to get sequenced while with the synthetic biology, which [00:30:00] is basically, you know, in this case is basically pharmacology. People are already consuming drugs and you're just making better and better drugs for diseases for which there were no drugs.

[00:30:09] And it's just a, it's an easier pipeline. While I would say that our first sequencing innovation was in 1984. And here we are in almost 2024. And we

still don't have, you know, 40 years later, we still don't really have consensus enough that health care providers think that it's a good thing to give everybody their whole genome sequence, uh, or, or act on it in any way.

[00:30:32] And most people don't do it on their own. So, there's a disconnect. I think it's the 1% dilemma. It's the seatbelts, smoking, global warming, and getting your genome sequenced. They're all like, hey, I got a 99% chance of doing okay. If I went to Las Vegas with those odds, I'd be fine. But it's different for public health.

[00:30:56] And it takes special effort to get people to stop smoking and to [00:31:00] wear seatbelts and so forth. So, I think that's what's going on here. And no, no government agency has stepped forward to do what they did for seatbelts and smoking, which was a whole, whole series of experiments. Like just passing a law to buckle your seatbelt wasn't enough.

[00:31:15] Getting them present in every car was not enough. They had to actually mandate a circuit that would close once you buckled it. On top of your belly, not underneath your, you know, that hasn't happened yet for, for Joe. Even though the, the carrier status alone, plus adult-onset diseases could save us a trillion dollars a year and a lot of pain and suffering.

[00:31:39] It just hasn't happened yet. But with synthetic biology, totally different thing. There's the orphan drug act. That makes it very profitable to go after rare things, and then there's lots of common diseases as well that can be treated. And my favorite treatment is gene therapy, for reasons that we could go into if you, if you want to.

[00:31:57] Raj, shall we go to the lightning round? [00:32:00] Sure, so George, we'd like to do a quick lightning round, if that's alright with you. We're gonna ask a series of questions, and the rules of the game are one to two sentence responses, uh, to each question. Yes or no is, is also great. Does that sound alright? And some of these will be highly entropic questions.

[00:32:20] And the goal is to learn more about George Church and how he thinks about the world, but you have to be brief. Um, so, uh, the first question is kind of a Turing test for biology is the way that I think about it. And an appropriate response to this could be that's a dumb question, but, will AI understand biology in any meaningful sense?

[00:32:41] Where understand is in air quotes here. I think in a way it already does. Highly advanced biotechnologists talk to each other essentially in

biotechnology, which is not really natural language. And computers, I gave an example for protein [00:33:00] design. Alright. George, what's your favorite piece of music?

[00:33:06] Favorite piece of music... Gee, you know, I kind of like Talking Heads. Oh, nice. You know, "this is not my beautiful house." I don't know. Given a little more time, I could go with a few others. Yeah, that's the point of this though, is to inject some entropy. So, um, uh, you've had a storied career in science, um, what one thing has changed the most either from a technology standpoint, from a society standpoint, from a political standpoint in science over the course of your career?

[00:33:40] Well, I would say all of the above in the Genome Project, because it, the NIH was entirely hypothesis driven. Now, you've got one institute that is discovery driven and another one that's engineering driven, and that coincided roughly with a twofold increase in the NIH budget while we were [00:34:00] starting the Genome Project.

[00:34:01] I can't say it was cause and effect, but it was, it was a nice coincidence. Do biologists need to understand machine learning to contribute to machine learning projects? Do citizens have to understand GPS and atomic clocks in order to find directions on Google Maps? Alright. We'll accept it. One sentence.

[00:34:24] We accept it. As we know, the price of things in health care tends to be sticky, so the question is, will machine learning ultimately reduce costs for diagnostics and drugs? Yes, is the short answer, uh, but I don't, and I'll give an example where it's not sticky depending on how you define things. So, gene therapy was 2.8

[00:34:47] million dollars a dose. Until we got to COVID-19 and the top five vaccines were all formulated as gene therapy, some as low as $2 a dose, so 2.8 to 2, so [00:35:00] give me, you know, that's not very sticky. Okay. Are preprints a net scientific good? Preprints, I think net is the key word, you know, they're, they're good and bad, pretty high levels of good and bad, but yeah, I think they're net positive.

[00:35:17] Okay. Final question of the lightning round. If you could have dinner with one person dead or alive, who would George Church have dinner with? Oh, geez, I think probably Nettie Stevens, you've probably never heard of her, but she was on the little Google logo, but I knew her before that. In around

1910, 1915, she found the chromosome theory of inheritance along with Morgan, but uh, or separate from Morgan, about the same time.

[00:35:45] Alright. Alright, so, we're going to move to the final segment of the episode. We're going to talk about some big picture things. I think it's fair to say we've already touched on some big picture topic so far, but we're going to try and broaden the aperture just a little bit further. [00:36:00] We've talked a lot about biotechnology and your work in the area.

[00:36:03] I want to come back a little bit to a clinical focus and given sort of what you see happening either in diagnostics or gene therapy, what medical specialties do you think are most likely to be changed and impacted by AI? Medical specialties, hopefully genetics, interpreting the genome is increasingly engaging polygenic risk scores, and I think that probably could be done better, and then that could have impact on almost every field of medicine.

[00:36:35] The other thing is age-related diseases. I think there there's an opportunity of having multiple genes involved in gene therapy and possibly even personalized or personalized medicine in general, but especially related to aging because aging affects every disease. Basically, almost every form of human morbidity and mortality is impacted.

[00:36:58] So, I think those are [00:37:00] a cluster of three things that interact with one another. Genetics, aging, and machine learning. So, thinking about our listeners who are clinicians and in particular early career clinicians, med students, residents, what do you think those folks should know about AI to help them prepare for a career in medicine?

[00:37:22] I think that they can have a fairly high-level view of it. It's like we no longer, most of us don't program in zeros and ones. We program with high-level languages like Python or maybe even HTML or Excel or something. So, it'll be like that. Hopefully, it'll be very easy to interface with. This is the case for most really awesome software.

[00:37:45] But they will have to know it. And they might not have to memorize as much. When I was a boy, you know, we had to memorize all the biochemical pathways and all the pathologies. And hopefully it'll be like, how do you look for it? How do you interface [00:38:00] with the machine learning and the big databases? So, you can't know it

[00:38:04] all anymore, but how do you know where to look for the answer? There's not going to be a Krebs cycle of AI for, for physicians to memorize.

Hopefully not. On the contrary, it's going to mean fewer people learning Krebs cycle. Excellent. And we'll learn instead, oh, uh, isocitrate dehydrogenase is very impactful on certain gliomas, right?

[00:38:31] And so it's one of the most treatable of the, of the otherwise nasty category of cancer that hits the brain. Yeah, I'm married to a clinician and I think that the mandate that I've been given is that if I develop anything in the AI space that makes her job more difficult, if I have a new Krebs cycle that she has to memorize, then that thing is not going to get very far clinically.

[00:38:53] I saw it, yeah. Yeah, there's, there's in this day and age there's really no [00:39:00] excuse for poor user interface in, in computing. Or another checkbox that you have to click, or something like that, exactly. That's true, there's plenty of bad software nevertheless, but there's no excuse for it, yeah. So, I'm glad that I get to ask this question, because I'm sure that we'll get a great answer.

[00:39:20] What is your most controversial opinion? My most controversial opinions were placed on me, not from me. Okay. So, like things having to do with advocate. I don't have to precede this, but I don't advocate, but some people felt that I was advocating cloning human Neanderthals. So, uh, a controversial opinion that is yours, not attributed to you.

[00:39:47] Or this line is probably that everybody should seriously consider getting their genome sequence. In particular, if they're of reproductive age. You know, let's say 16 and [00:40:00] up, especially for men that keep going, that they should know their carrier status. Uh, and that could influence who they date or various other things.

[00:40:09] I think the idea of a dating app that is aware of your carrier status is the most humane place to do it. But the controversy is they either think that that's eugenics, which it isn't, or it's controversial because you don't want to de-romanticize something by being so technical. But anyway, I think that's a huge missed opportunity.

[00:40:33] It's more humane because I'm glad we're not in the lightning round. It's more humane because if you do it after you're pregnant, then you have tough decision about termination of pregnancy, which is tough for essentially everybody where you're pro-choice or pro-life. And if you do it after you're married, then you've got the tough decision.

[00:40:56] Are you going to have children with this person, which means you're going to [00:41:00] do in vitro fertilization, which is no walk in the park. And so that's bad news. But if you do it before you've even met the person, then it means that out of a thousand people you could date, you're gonna, you know, date 990 of them, and at most.

[00:41:15] And so you can eliminate a few, and there's no false positive. problems at that point. There's definitely a false positive problem if you're doing IVF or, uh, termination or, or worse yet, you know, doing a surgery to remove organs that might be at risk for cancer. But there's essentially no false positive problem when you're rejecting 3% of the potential suitors. And at the risk of misattributing another quote to you, I think I've heard you talk about this before and it would go something like, there's a dating app, like you said, and silently behind the scenes, you're getting screened out from people who have the same carrier status as you so that you're never matched with someone who would, so that two recessive genes would, would come together.

[00:41:59] Right? So you'd, [00:42:00] you'd kind of not even know, that it was going on behind the scenes. You would just be matched with people that you would not have this problem with. That's right. That's right. That's accurate and, and that's humane in addition to all the things I mentioned. That's humane because another awkward time is you've decided to marry somebody and then you get the score and then you decide not to do this. Or if that matter you get people find out, you know, you're all ready to get married and then the marriage is off. And then everybody knows that both of you are carriers and so, in a certain sense, maybe in less accepting parts of society, you both get branded as why should anybody date them?

[00:42:38] And in fact, everybody, you know, 97% of people should date them, just not the 3% that are mismatched. And so, by avoiding anybody knowing that your carrier status, I think it's the most humane thing. So, both more humane and no scarlet letter, no stigma, like none of, none of those issues. Yeah.

[00:42:59] Right, so I think [00:43:00] that's the time to do it, or we could destigmatize everything, but that's, I think that's harder. It's hard to say which is harder, but it could be harder, yeah. George, are there any examples of a, I guess, I think you called it a dating app. Are there any examples of that in place, or is this, is just an idea?

[00:43:15] There are. Not exactly that, not an app, I mean it's not quite, but it's like that, which is Doria Sharim, which was started by a rabbi, I think mid-80s.

So, it's been around for a while, because he had, I think, four of his children had Tay-Sachs, which is a serious disease, kills kids painfully at age four-ish.

[00:43:37] And he just decided that there should at least be the option among his congregation. It's anybody that could have similar afflictions. And it's scaled up to, I think, eight or nine genes, typically, that are enriched in the Ashkenazi population. But in a certain sense, we're all at risk for those eight or nine genes and about a thousand [00:44:00] more.

[00:44:00] And it's not clear why it hasn't spread. It's been very successful in the populations that use it. where it's lowered the risk of such births by at least a factor of 10, and why it hasn't spread to other populations is, I don't think it's because one population knows more or less science than the other ones.

[00:44:20] It's something else. It's not that one population has necessarily more genetic diseases. It is true that some inbred populations have slightly higher, but that's not – the point is, we're all at least a 3% risk. So, our final question for you, George, and you can take this in multiple possible ways, it's up to you.

[00:44:41] What applications of AI to biology keep you up at night? Oh, yeah. Well, first of all, I'm genetically narcoleptic, and so nothing keeps me up at night. It's about 30 seconds is the median time to falling asleep. But what keeps me [00:45:00] up during the day, is anything involving discrimination. So artificial intelligence could more.

[00:45:08] So it's an interesting question. When we worry about discrimination, are we worried about it being too inaccurate? In other words, we're, we're stereotyping an entire people category of people that maybe have a a priori a low probability of living up to the stereotype. Or are we worried it's too accurate?

[00:45:28] Is it, or are we worried that it's not accurate enough or it's too accurate? And I think it's case by case. But in any case, it could have enough imprimatur of accuracy that it would be used, but still inaccurate enough that it could be abused. So that's one scenario. The other scenario is they could use it.

[00:45:46] You know, make personalized weapons, you know, once, once the terminators come, then our human failures could be in a certain sense public to the machines, even though they're not public. In other words, I can't [00:46:00] tell it, but they could figure it out. I guess a follow-up question to that is I know like your work in gene editing, you spend a lot of time thinking about how when

this technology becomes democratized and you can buy like a, a reagent kit for $10.

[00:46:15] Like how do we, what are a set of ethics and what are a set of protocols that we can use in a world like that? Is that at all similar to how you think about what's happening with AI either generally or AI in biology because a model that cost 10 million to create today will, you know, cost $10 to create five years from now.

[00:46:33] So, how do we think about these powerful technologies that are also being democratized at like a very, very quick pace? How can we sort of balance safety and progress in that kind of world? Well, we've demonstrated that we, that there's no such thing as a slippery slope. That is to say, there are documented cases where we were able to keep ourselves off the slippery slope and other [00:47:00] cases where we are, we're not, at least not for the whole population.

[00:47:04] There's always some percentage of the population that falls into the trap. So, for example, speed limits, there is no magic. There's a point where suddenly it becomes unsafe, but people tend to stay pretty close to the speed limits. Yeah, I guess the question was about either professional societal norms, so I know that there's been a lot of this in gene editing where there are groups that meet to discuss safety and come up with regulation.

[00:47:31] Is there any of that lessons from that community that transport to AI generally or AI in biology? Right, so, you don't need special groups to monitor gene editing. Uh, you have the FDA. And the FDA is very effective at keeping us bringing things out that are safe and effective. That also applies to medical devices, so it is possible the AI would fall in that [00:48:00] category.

[00:48:01] But it's also possible you can evade the category by making something that's not recognizable. For example, dating apps that don't seem to be regulated by the FDA, even though they could have a trillion-dollar impact on medicine. In a certain sense, they're not medicine. But, nevertheless, so even though people try to make regulations on top of regulations, so they wanted to have a moratorium on gene editing on top of the FDA, which has a moratorium on all new drugs, I thought that was a little crazy, redundant, and it didn't really happen.

[00:48:35] But with AI, if it does slip between the cracks, then there should be some kind of safety. Now, our track record for that is not so great in computing.

If you look at the Internet, there was very little of the foresight that existed, very little of it made it in, in time. So, there's wide open doors for hacking, for

[00:48:54] computer viruses, for identity theft, for abuse of children, and [00:49:00] pornography, and so forth. So, we didn't do such a great job there, and I hope we do a better job with AI. Partly because of fantastic educational media, by which I mean the Terminator. Alright, uh, maybe we'll edit it so that we don't end on a dour note like that, but I think that, uh, Yeah, that would be better.

[00:49:24] Ending on the Terminator, great. But I can't remember the last time I've had a conversation where we discussed resuscitating woolly mammoths and dating apps within the same hour, so it's been, it's been a really special conversation. Thank you

[00:49:36] so much George for being on *AI Grand Rounds*. It's been a pleasure. Thank you. It was great.

[00:49:40] I look forward to hearing it. Yeah. Great, thanks.