



Futures in Biotech, 25: From the Human Genome Project to Space Exploration with Dr. George Church (Part I)

Leo Laporte

Bandwidth for the Futures in Biotech is provided by Cachefly at cachefly.com.

Marc Pelletier

This is Futures in Biotech, episode 25 for Wednesday October 3, 2007. From the Human Genome Project to Space Exploration with Dr. George Church.

Futures in Biotech is brought to you by Invitrogen, the Zero Blunt TOPO PCR cloning kit. It combines a 5 minute TOPO cloning with a unique zero background technology. Visit invitrogen.com/topo for details.

[Music]

When Dr. Church agreed to do this show I knew we'd be talking extreme biotech, but I really – I mean really, had no idea. And since we only scratched the surface here this will be a Part I of two. So onto the interview.

You did some work on the Human Genome Project, pretty much at the epicenter of it in the '80s.

Dr. George Church

Yes, so maybe going a tiny bit further back, at one point I typed in all of the nucleic acid sequences. So I was a teenage crystallographer at the time, I typed them all in. I thought this is really great, we should really sequence everybody that wants it, and so I joined Wally Gilbert's lab and we worked out genomic sequencing and multiplex sequencing together.

And then around the time that – actually the first paper on genomic sequencing was 1984 and I went to this meeting at that the DOE held at Alta whose point was to get at mutations in people exposed to energy sources of various sorts, and chemical. And we decided in the first 10 minutes we couldn't do that, but hey maybe as an alternative we could suggest the genome project. The DOE liked that part and I was one of the first grantees in 1987. And then pretty soon NIH decided that the DOE couldn't have such an incredible chunk of – such a possibly interesting chunk of human health so they started a genome project.

And I helped three of the first centers get started, including one at Stanford and one in Waltham that became the only commercial part of the public genome project, which is called Genome Therapeutics, eventually. And then I helped, I was a co-PI on Eric Lander's first grant which grew into the Whitehead Genome Center and later into the Broad Institute.

Marc Pelletier

So one of the methods that you personally helped work out was sequencing straight from the genome.

Dr. George Church

Right.

Marc Pelletier

How did you go about doing that?

Dr. George Church

Well, so we felt that one of the interesting things to be able to do was to be able to quickly resequence. And at the time this was pre PCR, the only way you could do that was by making a library, screening the library and pulling like one clone out of this big library, and it was a whole production. And so we came up with the first, the pre PCR way of sequencing directly from the genome by basically sequencing the genome in its entirety at a chemical level, not reading it out, and then running it out – at the time everything was gel electrophoresis, and then probing it with one probe out of the genome, the way you would do a Southern Blot. And that actually worked pretty well and people wrote books on it and variations on that method were used for 13 years, including the first complete microbial genome. The first genome that was sold commercially in 1994, which was a year ahead of the first academically published genome in 1995.

Marc Pelletier

So what was that like, to be – well first of all this is almost Orwellian right, this is a change in era back in the early '80s. I guess PCs are starting to come out. The software to help us analyze these genome as they're coming out is starting to be made available. And you're actually looking at the code for the first time, not one gene at a time, but the code as it was written I suppose, or as it was evolved.

Dr. George Church

Well to me it wasn't that revolutionary in that it was more incremental. You're kind of close up to it, and I had – like I said I had typed in all the sequences in 1975 and so those 9 years was kind of slowly seeing more and more sequence. I had done the first plasmid sequence with Greg Sutcliffe and then we did one of the first operons and then the first complete genome. And it at no point seemed that remarkable. But certainly seeing the first complete genome we were one of the few labs that actually was unintimidated by the computational aspects of it and there was just a lot to do. There was a lot of details and a lot of integration of tools and thinking about the biology. And it was fun, it was definitely fun to think holistically about the whole genome and knowing that nothing was missing, unless you had messed up your software.

Marc Pelletier

[6:45] Did you have an idea of gene structure when this first started and what was known in the '70s and '80s in terms of how the genome was organised?

Dr. George Church

You know I think there actually were – now and then we'd describe a bunch of things as surprises or discoveries but I think there were really relatively few surprises. We had – the idea that you would have an intron was new, and I remember I was excited enough about that that I went to the seminar twice. But for the most part we expected genes to either have full regulatory sequences upstream from them, maybe it was a little surprising when we started finding them in the middle and way upstream, like enhancers. We expected mammalian, or a variety of genomes to have a lot of extra DNA that would take us a while to figure out what it all was, but we sort of expected it to be of structural and regulatory significance.

Marc Pelletier

So I guess you weren't – it didn't give you the sense that you were sequencing alien DNA, that really this was how we worked, and we've been working on it for a while and understood some of the components.

Dr. George Church

Yeah. I remember thinking – I thought overlapping genes were really cool because it was like the whole, how would you code that in the genetic code and all that. But then I had the suspicion that it was not going to be very general and it turned out it seems like it's not that general. There're very little protein coding regions that overlap.

Marc Pelletier

Are those mostly in yeasts and in bacteria where this happens?

Dr. George Church

They're mostly in icosahedral viruses, meaning where you have very limited amount of space for your genome.

Marc Pelletier

So this means that you have a gene that's encoded on the same strip of DNA. Are they necessarily in the same direction on the DNA, or can they be...?

Dr. George Church

No, they can be – there are 6 reading frames, 3 on the top and 3 on the bottom. And you can have two frames used in any combination.

Marc Pelletier

So this is like a crossword of sorts, that phenomenon.

Dr. George Church

Yeah but it turned out it was a pretty limited – very few organisms show this and it's fairly limited. But it was kind of cool. It was one of the – one of the first genomes display it, which was phi-X174.

Marc Pelletier

Well it certainly shows the power of evolution to conserve the amount of chemistry required to synthesize a genome and make as much use of it as possible, which is really neat. I guess, when you've been playing with genomes for 20 some odd years you sometimes have to take a step back say well, these are codes that regulate the process of life and...

Dr. George Church

[9:40] And that was the other thing that was kind of interesting but Fred Sanger found the overlapping genes and he also found new genetic codes. And I remember that that made me pause for a while and think about – there's really no dogma that's worth embracing so much. So in part – almost all these things had some surprise and in a way you say I'm not surprised but now there are quite a number of exceptions to the genetic code. And one of the things we do in the lab is we're trying to create new genetic codes.

Marc Pelletier

Can that be done? Can you reengineer...?

Dr. George Church

Oh yeah, sure. There's really – there probably aren't that many limitations on the genetic code, once you get into the synthetic biology way of thinking.

Marc Pelletier

So your work reflects both a systemic and a synthetic level of detail. Getting an understanding of entire genomes, to compare them, to determine data – or to retrieve data on biological processes, is that correct to say? And then you also try to reinvent or reengineer for the purpose of synthetic biology, repurpose the tools that we have?

Dr. George Church

Right. I mean, we try to not accept things the way they are. If there's a type of information that we need that isn't available with current instruments then we'll make a new instrument or a new software or wetware. If – and if there's something where we think we could improve on organisms with respect to a particular task, not their original task, but retasking them then that's kind of an extension of what we would do with instruments. With due respect for the ecological and other responsibilities you have when you're engineering organisms.

Marc Pelletier

So do you consider some of these organisms, as we discussed before we aired, as equipment?

Dr. George Church

[11:44] Well I think there's the debate as to whether DNA is hardware or software or some interesting combination. In a technological sense they are devices that have a special property of replicating, so they require a special kind of respect, which is why people get regulatory concerns about genetically engineered, genetically modified organisms. And that's why one of the things that we set as our goals for my laboratory and for the field of synthetic biology, which distinguishes it from previous fields, one of many things that distinguishes it, is trying to put in a lot more quality control and safety features, which is the way an engineer thinks about things. And I think that's what's – it's ironic that genetic engineering was really not so much an engineering discipline while synthetic biology is much more so. Not just quality and safety but hierarchical assembly, inheritance of modular constructs, specification sheets and all that sort of thing that you expect of an engineering discipline. That's what we're trying to do in the field of synthetic biology.

Marc Pelletier

Synthetic biology in a sense, our understanding of the function of some – I guess we have a great understanding of some proteins and very little understanding of other proteins. And – I don't know, at what point are we in our knowledge base of genes and their function to be able to really start constructing organisms with a purpose?

Dr. George Church

Well that's an interesting viewpoint. The analogy I would make is we understand some elements of the periodic table much better than others, we're not really in control of the transition metals the way we are in control of carbon, nitrogen, oxygen. And so we tend to as chemists and electrical engineers use a very small set of elements that we trust. So to make a large-scale integrated circuit we might use silicon and boron and phosphorous and copper and gold and a very small number of the 100+ elements. And the same thing with synthetic biology, we are not obliged to use or even understand most of biology, we simply focus on the parts that we trust. That's the key thing. And then we take the parts we trust and by using them over and over, and debugging with them and so forth we learn more and more about them, we trust them more. And we alter them so that they're more trustworthy and we put them into constellations where we can specify their parameters more precisely. We can move away from the random behaviour, the unpredictability of parts that we don't understand into making them much more usable as tools. So it's really a subset, as your question prompted.

Marc Pelletier

How large a subset? So we tend to think of the -omes, proteomes, genomes, the transcriptomes and exomes or whatnot – ORFeomes, looking at entire sets. But if you're saying we don't need to necessarily understand all of them, but understand the variability I guess or make predictions as to how reliable we can get things to happen, how big a part of the -omes do we need to get a grasp of to really take advantage of playing around with these processes?

Dr. George Church

[15:36] Well you can imagine that you could take a fairly tiny fraction and expand it. So for example let's say you really understand zinc finger DNA binding proteins much better than any other DNA binding protein, you could just forget about the others for a while and expand the zinc fingers. So you could start with one zinc finger and make a whole library where you have every possible sequence there. And to some extent that's the plan of the company Sangamo with which I used – I helped in early days. And that gets you a lot.

Or you could say I don't completely understand RNAi but I can make riboregulators from first principles. And that's the sort of thing that Christina Smolke and Farren Isaacs in my lab did –

made riboregulators pretty much from first principles. Or you can select them from scratch where you make completely random polymers and bind them to something you want to have them bind to, and the tiny subset of that set of random polymers that binds you can amplify. And now this is a part – so you made this part basically from scratch, based on functionality, what you wanted it to do, which is an opportunity you rarely get in most other fields of engineering where you just slam a bunch of random stuff and the right thing comes out.

So all that required in terms of understanding was we needed to understand something basic about RNA is that they could fold and that evolution works in vitro from a completely formless population to something that has functionality. So the part was almost entirely from scratch. So that's 2 examples, an RNA example and a DNA-protein interaction example, but you could go from there. You really only need one example of each functionality.

Now when you get to enzymes you probably want lots of enzymes because those are hard to make from scratch, either by design or by evolution, like the aptamer or the binding RNA. And there you might actually want every – at some point a collection of every enzyme that's every done any reaction that you're interested in, even if it requires getting them from a lot of different organisms. Which is another thing that synthetic biology does very well is you gather up whole pathways from multiple organisms and put them all in one organism.

Marc Pelletier

Are projects like the Venter around the world important in building those tools, or at least getting some of those genes that are novel for those pathways?

Dr. George Church

That's a very interesting question. I fully support that kind of project but take the zinc fingers example, in that case you needed one well characterised zinc finger and then you – and you had the crystal structure and you went in there and you changed the nucleotides that matter to the DNA interface and you generate a whole library of every possible zinc finger you could ever want. And then you can select for them in vitro to do what they want. Now, that said I also said that for enzymes we want every enzyme in the book, because these exquisite detailed – they're like watchmaker type things. Zinc finger you can think of as kind of like clay binding to a surface, it's really –

Marc Pelletier

You can model it.

Dr. George Church

[19:14] Yeah, it's [surface to surface], so it's easy. But with [enzymes there you might want to get a lot. But the problem with the – going through the ocean is, you only see things you already know, so you don't really get – easily get new enzymes. Or you get the new enzymes but you don't know what they do. So you can get lots of examples of a hydrogenase but you already have two really good examples of a hydrogenase, the nickel-iron and the iron-iron hydrogenase. And the other, all – they fit in those two categories. But you see lots of solutions and maybe if you want to perfect a parameter you'll take a lot of different genes that you got from an environmental survey and you'll recombine them with one another in vitro and if you have a selection of a screen for that process then you'll find it, either in the original collection you found in the environment or recombinants you've created in the laboratory. So there's some power there. It hasn't been demonstrated in that many case but I think this is definitely worth looking at.

Marc Pelletier

If I draw a parallel to the chemistry that's found in the rainforest, I suppose even though we could probably synthesize most of the organic molecules from scratch, seeing them in their environment gives you the context. In which case you know that you need a tool to create an enzyme that can bind carbon dioxide under very stressful conditions or whatnot, or identify changes in pH in the water, you need those environmental cues I suppose to really – I mean the

chemistry in the rainforest, we could randomly synthesize it but a million monkeys typing – you know what I mean?

Dr. George Church

Yeah I think that's a fairly good example although it's more like the aptamer, the binding example because the chemicals that most drug companies get from the rainforest aren't necessarily doing in the drug company what they do in the wild. In other words we basically don't know what their function is in the wild, and don't care. So quinine may not have been invented by the plant to cure malaria, probably wasn't. It had some other original purpose. And so we basically collect up all these chemicals and just grow them against problems and see which ones stick. And I think that's more like the aptamer.

But the exquisite enzymes, if we found a new enzyme that did a really cool new thing, like say made carbon nanotubes in the rainforest or in the ocean, we wouldn't know it. It would be sitting there waiting for us to build cables into space and we wouldn't know it, unless somebody had the insight to try a lot of assays.

Marc Pelletier

So synthetic biology then to me you'd need – or it would benefit from having an organisational structure of having a lot of effort on retooling what's known and then discovery science going on looking for amazing things or amazing new tools.

Dr. George Church

[22:45] I agree. I think there are certain things the biology community kind of has a parasitic mode where it waits on many other fields to make the discoveries of a new class. And then the synthetic biology acts as engineers to take that class and expand it and characterize it and fit it in so it's interoperable with other engineered parts. The discovery mode itself is not something that the synthetic biology community has currently mastered – feels like they have the master of, it's something that they more get from other fields.

And I think the metagenomics you talked about, going through the rainforest or the ocean doesn't do – needs to and occasionally does functional screens where it's not just doing sequence screens but actually trying to find new functions. But I think there's a possibility for those fusing in the future where synthetic biology allows us to screen for lots more functionalities, like creating really elaborate selections rather than simple selections like what binds to this, or is resistant to this. You could actually make a very complicated – you'd make a complicated device to look for a complicated device.

Marc Pelletier

Do you view synthetic biology as a method to create entirely new organisms for – do you feel that is best used in health, in the field of health where retooling might only require one or two genes to be changed? Or using for health in changing someone's physiology entirely by complex traits, by multiple genes?

Dr. George Church

Well there's a bunch of questions there. Synthetic biology is not just limited to health certainly. There's biofuels; there's all kinds of chemicals that it can make; antibodies; in vitro and in vivo biosensors. But –

Marc Pelletier

Terrafication?

Dr. George Church

Excuse me?

Marc Pelletier

Terrafication. So last week we had the Phoenix Mars Mission and they're going off onto Mars to look for the conditions where, one, organic molecules are, if life was there. And/or determining the resources required to terrafy.

Dr. George Church

Well, terraform I hope, not "terrify". What I'm worried about is we terraform it before we discover whether there was life there. And if there's fragile life there, or any kind, I think we should try our best to discover it before we contaminate it. Because it could be mind-boggling.

Marc Pelletier

Well they have some phenomenal instrumentation that would – if there's going to be life the bets are they would find it.

Dr. George Church

Well we have a NASA funded project to send a miniaturized PCR machine up to Mars. The principle being that there's a lot of exchange of material between Mars and Earth, and if life evolved on Mars and came to Earth, or somehow made it from Earth to Mars then they should have a common ancestry. And this could be a really unique, could be a very ancient separation, and so that's worth finding. And it's also, if there is a little ecosystem there, say in the poles where there's more water or deep in the earth, we should understand that because it's probably the most alien environment that we're going to see and it would profoundly change the way we think about our place in the solar system.

Marc Pelletier

[26:51] So how did that project get started? How does one decide "hey, I think that if there is co-evolution, or the origin of the primordial soup had it been on Mars and here, and exchanged"? How do you put that into the form of a grant? I'm just curious, that's an absolutely phenomenal hypothesis to test.

Dr. George Church

Yeah, well...

Marc Pelletier

Do you talk about it first with the people at NIH or NASA?

Dr. George Church

Well, I don't think NIH would fund that. We did talk – this was kind of a conversation that came out of conversations that Gary Ruvkun and Mike Finney and I had, Maria Zuber at MIT, and we put together with pennies that we had lying around and many years without funding. And then eventually qualified for – NASA approved the project but didn't fund it, and then later it funded it. And then there's this whole series of steps you need to go through to prove, like two years in advance, that your method is really working and then there's another two year delay before it actually is on a mission. There is a window of opportunity every two years because of the relative planetary motion.

Marc Pelletier

This is genetics on a planetary level, this is phenomenal!

Dr. George Church

Yeah, it's a fun project. I mean it's a bit of a long shot but we do know that rocks are exchanged between the two planets.

Marc Pelletier

Do you need proof that there's organic molecules on Mars before you send a probe out there with a thermocycler to do PCR? The audience, just to remind them, at least my dad who's listening – PCR is a method, polymerase chain reaction, that allows you to take one single molecule, or just

about one single molecule of DNA and then make trillions of copies. So if there is one copy left you can recover it and amplify those genes, I guess somewhat in the same way that Jurassic Park would amplify or clone a genome, you can bring things back. And you actually have done most of the work, or a lot of the work on some of that single genome copy – cloning by PCR cloning?

Dr. George Church

Right, yeah polymerase colonies or polonies, or polymerase clones or plones. It's a great in vitro way of getting nucleic acids amplified from as little as a single molecule. Or even a piece of a DNA molecule, which has multiple atoms.

Marc Pelletier

So the question was do you need organic chemistry, or do you need to know that there's the chemistry on Mars to be able to send that probe?

Dr. George Church

[32:05] I think if we're searching for evidence of life on Mars we need to try multiple things simultaneously, because of the cost of these missions and the infrequency with which we can go there. So I think it may turn out that it's easier to detect DNA than it is to detect organic molecules, because DNA can work with very trace, single molecules in fairly large volumes, while most analytic methods require millions of molecules per microlitre. Or per reasonable droplet that you would pull out of the soil, so there's an exquisite sensitivity to PCR and you can also think of it, if you do find organics then you might doubt it. But if you find organics and you find PCR on the same mission then suddenly you've got a story that's much more believable.

For example in the first Mars mission that tried to look for life they said in advance the criterion by which – if we see this then we will declare there's life on Mars. And lo and behold, they did the experiment and they saw – it passed the test. And what was their conclusion? They said, it's a badly designed experiment, but they basically backtracked, they made a decision – if we see this result then there's life on Mars. They saw the result and they didn't conclude what they said they were going to conclude. And if they thought it was a badly designed experiment they should have decided that in advance. Why does it suddenly occur to them just because they got positive results? So I think that's why you want to have multiple ways of seeing it and have less ways for people to slip out of the conclusion there is life.

Marc Pelletier

How would you eliminate the possibility of a false positive, or that that probe is carrying DNA from the JPL, or Jet Propulsion Laboratory members over to Mars. If it can be from a few atoms...

Dr. George Church

So what you hope for is that the sequence will be a sequence never seen before. And we can do all kinds of control experiments where we sample dust in the lab where the lander is assembled. Everything is sterilized on these missions because they don't want to introduce bacteria from earth anyway, even if we weren't doing PCR they would have very strict rules for... But nevertheless if we sequence, in some sense identify some of the sequences that we amplify on Mars and they don't correspond to anything on Earth, including all the controls we did, then that's fairly convincing I would think.

Marc Pelletier

Do you have any graduate students working on this too?

Dr. George Church

It's mostly professional staff, it's a fairly small project. It's distributed among our various labs.

Marc Pelletier

Because it would be a really fun thesis.

Dr. George Church

Certainly once the data starts rolling in it could be fun. It's fun in the engineering side as well. In a certain sense many of the things we do in the lab feed into, it's like all the projects in the lab are interdigitated in various ways. They synergize. So the more we learn about PCR the easier it is for us to miniaturize it so that it fits in less than a 1kg payload. And the more we learn about contamination in the lab the better we can protect against contamination. So in a way, none of the graduate students are and all of them are, depending how you look at it.

Marc Pelletier

I'll have one last question on the Mars project. It actually took me very much by surprise. But of course if you get a positive and it's really a positive, this is of course one of the biggest discoveries of the decade, if not century, if not since the first scientist that looked up through broken lenses and saw canals. Who was the astronomer that looked – Herschel?

Dr. George Church

I can't recall.

Marc Pelletier

I'll edit that part out. So what if, what if it wasn't DNA. And your PCR is based on the ability to amplify from DNA. Is there any chance you can send in a reverse transcriptase, because the theory is that life first started off as RNA or replicating molecules of ribonucleic acid before DNA. Do you see RNA as a possibility?

Dr. George Church

[37:10] Oh yeah, I think we certainly want to be able to get the broadest set of nucleic acids, maybe even things that are a little bit beyond RNA and DNA is possible – challenging. But certainly RNA and DNA are relatively straightforward to get those as something we can detect and analyze, sequence.

Marc Pelletier

How about before even the self-replicating molecule of RNA?

Dr. George Church

I think that moves into one of the other modules. For example there are some modules that are looking for any kind of chiral asymmetry, that is a most – a very general capability which I think is important to have on the mission. The original one was to see if you had any kind of metabolism. So there are a number of different approaches but we're focusing on nuclei acid-like polymers.

Marc Pelletier

When are you sending the box up?

Dr. George Church

I don't think the date has been set. Could be at the earliest 2011.

Marc Pelletier

Okay. Well I'm certainly glad you're doing this because this is going to really stimulate the world of science. It's like buying a lottery ticket to an extent, lottery type science where if you don't buy the ticket you're not going to win but – somebody's got to be trying.

Dr. George Church

Exactly, exactly. And we certainly need to know – even if we're not planning on terraforming, if we're just planning on sending people up there, or even complicated robots that could have a higher chance of contamination, the sooner we figure out what's there already the sooner we avert an ecological loss. Lost opportunity. So I think it's a fairly high priority to rule out life if we

don't find it. If we need to very rigorously rule it out then by sending probes to very many different locations and drilling down and all that...

Marc Pelletier

Well even if there was just traces of organic molecules that were left, sort of footprints of...

Dr. George Church

Well the problem is there are organic molecules throughout interstellar dust. So merely having carbon covalently bonded to other atoms isn't enough.

Marc Pelletier

Well maybe large complex metabolites I suppose, traditional metabolites or maybe less traditional metabolites. I think if there's anything positive the world...

Dr. George Church

Well there was something positive in the first mission. We have a much higher bar now. The chirality is something you don't get from – you don't expect from interstellar dust without life. Although you find some –

Marc Pelletier

Can you define what chirality is?

Dr. George Church

So chirality is if you have a collection of atoms they will have a handedness to them. So let's say you have four atoms arranged in a tetrahedron, four around a central fifth atom. Then you can arrange them into mirror images and if you see a preponderance of one mirror image, ideally an exclusive one. So almost all of the molecules in our bodies, a huge fraction of the ones that are polymers, like DNA, RNA and protein have a handedness to them. So if you tend to see one handedness and not the other then that's a very good indication of life.

Marc Pelletier

If all the compounds are left handed then you know that was an enzymatic...

Dr. George Church

That's right.

Marc Pelletier

Well keep an eye open for that. So that was already found?

Dr. George Church

No that was not. The original test was more of a metabolic test looking for evolution of carbon dioxide gas and they at some point, I think after the positive results came in they came up with some way to explain it away with ordinary chemistry not living chemistry. Which they should have thought of in advance because it sounds like a post hoc wimping out.

Marc Pelletier

You have to learn how to do an experiment and the only way to learn how to do an experiment is to do an experiment, I suppose. You never get it right the first time.

Dr. George Church

Well they could have visualized that outcome. They could have said, "oh yeah it produced carbon dioxide gas, now what are we going to say?" And they could have said, "we're going to say it was a bad experiment." They could have announced in the beginning that no matter how this turns out we're not going to declare there is life. They could have announced that. Maybe that wasn't dramatic enough.

Marc Pelletier

Maybe they need some better reviewers at the point of putting the project together. So actually we've got a lot of great material here but I was wondering, there's so many different areas of your work that I'd like to go into, maybe we could have you back in a couple of weeks or in a month or so.

Dr. George Church

Sure whatever you want.

Marc Pelletier

Because this would be really great. I think that this would be fun to do, if we can do a follow up episode. We did it with Marc Vidal and it really gave us a really good understanding of his interactome work, because it is fairly complex. And one area that I'd like to talk to you about is how you see genomics at the system level, how do you pull out information that's relevant to processes like aging and cancer – and so that's one area that I'd like to talk to you about. Another is I'd like to explore some more of the synthetic biology. So how about this as an exit question, and then we'll have you back. So in, was it Nature Genetics – yeah Nature Genetics the question of the year last year was, what would you do if it became possible to sequence the equivalent of a full human genome for only \$1000, what would you do? And what was your answer, do you mind answering it for us?

Dr. George Church

[43:45] Sure, my answer was that if you – if it really scales perfectly, which is possible, then you should be able to do a subset of the genome for less than \$1000, and maybe 1% would be \$10 or something like that. And to some extent no matter what the price is you might want to get more genomes for less money. And in any case right now you might be able to do 1% of the genome for \$1000 and get a million people sequenced and interpret that data in ways that they may find interesting.

Marc Pelletier

How would you pick the 1% that you want to sequence?

Dr. George Church

Well you would start out with an intention to get the whole thing but you would start out with the protein coding regions, splice junctions which are right next to them, a few extra percent for that – not percent of the total, a couple of extra base pairs for every splice junction. RNA start sites, one per gene, or more. And that might add up to 1%. And then you could make it 2% as you find more regulatory elements that are conserved or not conserved between organisms. And that would be a starting point. Anything that's shown to be a causative allele in a disease, or is likely to be a causative allele, then you prioritize those up. And there will be, there will probably be a 1% of the genome which accounts for 90% of the traits, we won't necessarily guess that 1% perfectly the first time but there will be a huge enrichment, and that enrichment is what cost effectiveness is all about. If you can get 95% of the information for 1% of the cost then you should do it.

Marc Pelletier

Is that happening? How many genomes have been sequenced right now? Is it two or three human genomes?

Dr. George Church

It depends on what you mean by sequenced. No human genome has been sequenced in its entirety, and about almost 300 have been done in part, meaning that you have a reference sequence and then you superimpose on that variability due to single nucleotide polymorphisms, insertion deletion elements both large and small, when they get very large they tend to be called copy number variants, and then various other rearrangements like inversions. So you have all these things that you can superimpose on the genome and the more of them you get the more you have characterized a new and individual genome. So now that we're doing genome wide

association studies with whole genome single nucleotide polymorphism chips, there are literally thousands of people are being analyzed as diploid, meaning both mom and dad's inheritance. That's sort of a 70, 80% level for single nucleotide polymorphisms, it doesn't cover all the indels [insertion/deletions] but in terms of things that are linked to disease this is the plan. But...

Marc Pelletier

So the efforts are focusing right now on just the single mutations or small mutations that are associated with disease on a large scale?

Dr. George Church

Well it turns out that a good fraction of the variation that you see among people is in single nucleotide polymorphisms that you can track with very inexpensive chips, about \$250 per person per assay – I mean per whole genome survey. And those, either the point mutations themselves, or something to which they are linked, which could be a more complicated insertion/deletion could be kilobases away. You could use those as surrogates so very cost effectively you can study – and we have now studied, thousands of individuals which are cases and controls with and without a disease. So the answer is nobody has been sequenced in their entirety and by this trick of having markers as surrogates for additional variation thousands of people have been done.

Marc Pelletier

[48:33] Wow. Do you see a time when the cost will get so low that everybody will just have everything, including the junk – what they call junk?

Dr. George Church

Yeah, we don't know if there's junk yet. But there certainly are regions that are much more neutral meaning it doesn't matter what happens there more so than other regions. It might get – it's been my goal for about 30 some years to get it affordable, sort of in the 1000, \$2000 range for whole genomics, but I think the way to get it there is to do a portion so that you can get it at \$1000 today and then instead of bringing the price down you bring the product that you get for that fixed price be more and more valuable, and upgraded every year.

Marc Pelletier

Sounds like you're very proactive because that's the kind of approach that will get work done now and won't wait.

Dr. George Church

Exactly. I'm very proactive, very dedicated to getting this available to as many people who want it. And also making it interpretable and socially responsible in a way so that it's not over-interpreted and misused.

Marc Pelletier

Well there's going to be a transition towards personalized medicine I suppose when you get that. I know, I personally have had a number of genes sequenced for some of the things – haematologists are doing it more and more and as the technologies get better and better they can focus on trying to get, eliminate misdiagnosis by finding out what people have inherited from their parents.

Dr. George Church

Yeah, I agree it's very important.

Marc Pelletier

It'll save a lot of lives. So I really appreciate you coming on.

Dr. George Church

Sure, my pleasure. Look forward to chatting some other time.

Marc Pelletier

Yeah well there's actually so much to cover. We'll definitely have to continue this, I've got so many more questions but I'll save it or else nobody's going to be able to download this, it'll be too large.

Dr. George Church

Okay well feel free to cut whatever you want or save it or what – okay?

Marc Pelletier

Okay, thanks a lot. Goodbye.

Dr. George Church

Bye bye.

Marc Pelletier

One thing that I've learned through doing these interviews is that some of the best ideas and thoughts come after the interview so what I did is kept the tape rolling and asked Dr. Church's permission to take the recording. So here we go again.

Dr. George Church

... people, it's kind of like – I think it's at the black hole stage now. It's like no one has ever used black holes to drive their car or anything, right? And for the most part the discoveries of the genome project were discoveries that only a mother could love. Like the number of genes and copy number variants, things like that. They either fell in the category where we sort of already knew it or it's a detail, or it's a million details.

Marc Pelletier

There's a lot of information that just from the genome itself we can medically – simple for medicine, not just understanding the biology of the human organism but for example I have a metabolic disease and it was mapped down to a gene. But unless my brother sequences that gene he's not going to know whether he has it and of course sequencing is the key. The answer is there and it lies before him and he just has to go to the right clinic that can do sequencing and sequence a specific gene.

Dr. George Church

Yeah I think people are going to be just – insatiable appetite for this. It's going to become like knowledge of internal combustion engines at the turn of the 20th century is that everybody's suddenly getting cars and you just got to understand it. Or computers in the 1990s when the Internet started getting accepted. Just everybody's going to learn about it and it's going to be insatiable. And it's going to be about them and unlike the Internet where you can say I don't need to worry about that because I don't need to have it in my house you've got your body, you're stuck with – you can't get rid of it and you've got to understand it and everybody else's.

Marc Pelletier

Well it might be the century of biotech. We are exploring other fields of science but it's undeniable the amount of information over the next 50 years is going to be tremendous. And the concept that we just may, with the combination of all the -ome projects, the proteome and transcriptome and interactome – the amount of information and then using it, apply it to synthetic biology as well we're coming to a point, what I call the singularity in biotech where our ability to control lifespan, metabolism all these things can be, either pharmacologically or genetically modified. And this is inevitable with all technological development, we're now just moving into a world of life technology. It's going to be fun, this is going to be great and I hope to cover it and give people a firsthand account...

Dr. George Church

[53:53] Singularities may not be fun. We don't know, we've never – very few people have been into a singularity and come back to talk about it. But overall it probably should be fun, at least until the actual singularity occurs.

Marc Pelletier

I'll refer back to the episode on the singularity with Drew Endy's answer to that, he had a brilliant answer and I'll let people tune back to that. He had so many different levels of analysis it was great. Okay, well thanks again. Do you mind if I put some of that on? A little trick in recording episodes is sometimes people feel more relaxed and there's some great content, great information so I'm going to...

Dr. George Church

As far I'm concerned I assume everything's always recorded, even when I'm not talking to somebody. So yeah you can use – I try to always say things that are suitable. And I'm not even that concerned about my speech impediments, you can put those on if you want to – we have to be open and transparent, right?

Marc Pelletier

Okay I'll let you go and I'll contact your assistants and we'll schedule the second part of this and talk a little bit more about applied genomics. Thanks a lot.

Dr. George Church

Thank you. Bye.

Marc Pelletier

I would really like to thank Dr. George Church for being on the show. It was really kind of him to accept the invitation. He's a Professor of Genetics and the Director of the Center for Computational Genetics at Harvard Medical School. I'd like to thank Phil Pelletier and Will Hall for the opening and closing themes. And last for the disclaimer, the ideas and opinions expressed here do not reflect those of Yale University or the Yale School of Medicine. Futures in Biotech, I'm Marc Pelletier.