[music]

0:00:06.3 Sarah Crespi: Welcome to the Science Podcast for February 5th, 2021, I'm Sarah Crespi. Each week, we feature the most interesting news and research published in Science and the sister journals. This week, we're dedicating the whole show to the 20th anniversary of the publication of the human genome.

0:00:22.2 SC: Today, millions have had their genome sequenced. This remarkable progress has brought with it issues of data sharing, privacy and inequality. I talk with three researchers about addressing these complications and the next step for genomic science.

0:00:41.7 SC: This week, we have a special issue on the anniversary of the publication of the human genome and its complicated legacy. Twenty years ago this month, we're talking February 2001, draft sequences of the human genome were published. Since that time, genome sequencing has become easier, cheaper, faster. Anyone can have their genome sequenced, but what can you do with that information? Who has access to it? Whose genomes are used in research?

0:01:10.2 SC: I spoke with a number of researchers about the state of genome science. We'll start with Yaniv Erlich, who talks about privacy in the age of easily obtainable genomes. Then we go to Charles Rotimi, who discusses diversity, or lack thereof, in the field, and what it means for the kinds of research that actually happens. And finally, Dorothy Roberts talks about the seemingly never-ending project of disentangling race and genetics.

[music]

0:01:43.0 SC: Now we have Yaniv Erlich, and we're going to talk about privacy issues around genome sequencing, but before we jump into that, could you just share a few reasons why it is good that so many people have their genome sequenced, something like 30 million people around the world. What are some good things about that?

0:02:02.3 Yaniv Erlich: We have people that are able to locate their lost family members. I know personal stories of people who are adoptees and their adoption certificates are just incomplete and they cannot trace their biological origin. So by sharing all this genomic information, we have so many success stories of people locating their lost relatives. And the second thing beyond that, by many people sharing their genomes and also phenotypic information, we can start to understand the genetic basis of virus conditions. A recent example is just understanding the genetic basis of being a severe COVID case which was published by a large consortium, just because people are sharing their data and also documenting their phenotypes.

0:02:45.1 YE: And maybe just one perfect point to highlight, the notion that we need to share genomic information really started around the Human Genome Project, with the Bermuda Principles and so on. This was something that the community really pressed this point, and also infrastructure was created to share genomic information. The same infrastructure and the same spirit, this is what gave us the ability to share the SARS-CoV-2 genome from researchers in China with researchers all over the world and enable to develop the vaccine in such record time.

0:03:19.9 SC: Now let's dive into some of the big issues with this, privacy first. So it's a lot harder to anonymize the genome than maybe was expected back when people first started sequencing them about 20 years ago?

0:03:31.5 YE: Exactly, yeah. We thought first that the genome is just very similar to any other biometric identifier, such as fingerprints or face structure. Now, it's very hard also to anonymize them, but the concept usually is that if your fingerprint is not in the database and someone tries to get your fingerprint from a coffee mug, they will be able to get your fingerprint, but they will not be able to put a name on the fingerprint. The genome is more complicated, because we share parts of our genome with relatives. These relatives share their genome around, they share this information, they publish it, and this enables people and adversaries to take an anonymized genome, search in these databases, find your relative and from your relative, there is a good chance that they can go and trace you back.

0:04:21.9 YE: Let me give an example. So we published that, actually, in Science eight years ago, we showed that we can look at the Y chromosome of males, and in some cases, we can infer their surnames by internet searches with some Y chromosome surname databases that citizen genetic genealogists just shared across the world. So we can go to the internet with these genetic markers of the Y chromosome of an anonymous genome and, in some cases, infer their surname, and after that, we can go and actually identify the person.

0:04:54.8 SC: And there's lots of cases where different agencies are using this to track people they suspect of crimes using genomes from close relatives.

0:05:03.8 YE: Yes, there is, of course, the famous case of the Golden State Killer, where the police had his DNA for decades, but it didn't show up in any police database. A few years ago, they consulted a genetic genealogist, and she suggested that they will infer genome-wide markers on his genome and will upload this information to a website called GEDmatch. They were able to locate a third cousin, then they employed genealogists that constructed a large family tree, and in the end, they were able to zoom on one profile that matched everything they knew about the Golden State Killer, then they swabbed the knob of his door, found a perfect match, and then arrested him, brought him to justice, and now he serves a life sentence in one of the California prisons.

0:05:50.7 SC: This kind of thing that we're talking about leads to the issue you raise in your piece about genomic surveillance and how this is something we can look forward to in the future, or not.

0:06:01.3 YE: Yes. [chuckle] There is some good aspects in it and some bad aspects, as every technology. The idea is that what the police did in the case of the Golden State Killer, they didn't have any special access to genomic database. Everyone with a computer and access to the genomic information can upload genomes to, let's say, GEDmatch or other websites, search for relatives, find a match and infer the identity of the genome of interest. So kind of the US is in the very unique situation where it's the first country in the world where you can deploy basically population-scale genomic surveillance. Another study we published in Science a few years ago showed that you can infer potentially more than 60% of those individuals with the databases at that time. Now the

databases are larger, so you can potentially infer more people. This means that we are in a very unique era where you can collect genomes, it's very cheap to develop the genomic information to get digital information from the genome, and then search in these databases and identify people, and this is why we call it ubiquitous genetic surveillance.

0:07:12.6 SC: You point out in your piece too that the harvesting of genomes or this genomic surveillance can be accelerated by the ongoing pandemic. How would that happen?

0:07:21.0 YE: We kind of like think that the public at this point is more willing to give some biological material, and that's a good thing, for sure, right? We need to test people in order to cut the transmission and to identify COVID cases. But if you have an apparatus where you can actually swab people, basically you can get also their DNA using the same swabs, that could lead also psychologically or sociologically that that's okay to be swabbed at airports, and therefore governments could use this information in the future also to collect the DNA of people and to create genetic databases. Now, I want to be very clear about it. I think it's a very good thing to swab people for infectious diseases at airports, but we also need to have discussions about what are the limits of collecting this biological material and making sure that it's not used for other things or if it's going to be used for other things, at least let's be transparent about it and let the public decide.

0:08:21.5 SC: What are some of the consequences, good and bad, for every single person in the country or even in larger areas of the world having their genome sequenced in a big database?

0:08:32.0 YE: We started our conversation saying that we can infer the genetic basis of different diseases and conditions, and that's amazing. We can offer new drugs, new treatments. It's important for families with rare genetic disorders that we do that. And they suffer horribly from these disorders, so there are so many great things about it. Also, our ability to catch criminals using these technologies, it's amazing. The day that the Golden State Killer was captured was a very good day for humanity. So if we have these large databases, presumably in the future, we could virtually solve any rape case. Nobody can rape someone and just hide and we'll not be able to find this person. On the other hand, we know from other countries, not the US, that sometimes genomic information can be used to repress minorities. Sometimes this surveillance, it's not a good thing, so it really depends on the political climate of the country. And this is why as scientists, we have the responsibility to alert the public about the different possibilities and what can go wrong with them.

0:09:38.0 SC: This is a special time. There are many, many genomes in sequences in databases that are pretty open, but I think even since the Golden State Killer case was closed, some of them have been shut off or have been disconnected from police activities. Do you see more guardrails being put in place either in the last decade or maybe in the coming decades?

0:10:00.3 YE: I think we start to see from policy makers putting some frameworks and policies how to use these databases. For instance, there was an interim policy by the Department of Justice, and this is great. And I think we had the conversation, we had a workshop in Cold Spring Harbor about a year and a half ago with people from the FBI, with people from the ACLU, with privacy experts and so on, so we all kind of like convened and talked about it. And you can see that these people coming from law enforcement, they understand that they just got some new super power that they didn't have, and they understand that this super power, there's some limitations. They cannot

utilize it too much. They cannot do bad things with it because at some point, the public might revolt and they will take this super power from them. So I was very pleased to see how deeply they think about it, and they tried in this policy to create kind of like some checks and balances. Now, it's imperfect, of course, it's very initial policy, but it was very good to see that they start planning ahead and thinking how to limit and just to utilize it for special cases.

0:11:10.2 SC: One other question, going back to privacy and research, is it possible to anonymize a genomic database for medical research? Is that something we could do if we wanted to?

0:11:21.8 YE: So I've thought about this problem for a long time, and I realized, I think it's like a kosher lobster. It's either kosher or either lobster but it cannot be both, right? So you cannot really anonymize a genome because you basically, if you try to do that, most likely you just lose all of the utility of this genetic information. Also another thing to consider, that privacy is not binary. There are 50 shades of gray in privacy, and there are different types of risks that we can take. I think the conversation should not be about how to protect the privacy of research participants but really to create trust relationships with them, because in many cases, privacy is not the end goal of people.

0:12:07.1 YE: People in many instances, they decide voluntarily to give away their privacy in a kind of like very limited way but... For instance, you go to your physician and you have to be naked in order for her to check you up, so you give some of your privacy for that, but you know it's for your own benefit. So if we can develop these trust relationships between researchers, between the general public, then we can really, I think, recruit individuals, like convince them to donate their genome and to see all the wonderful things that we can do with genomic information, and we can do really wonderful things for humanity.

0:12:45.1 SC: Thank you so much, Yaniv.

0:12:46.3 YE: Sure, my pleasure.

0:12:47.6 SC: Yaniv Erlich is an Associate Professor of Computer Science at IDC Herzliya and CEO of Eleven Biotherapeutics.

[music]

0:13:01.7 SC: Now we have Charles Rotimi, and we're going to talk about diversity in genomes and in genomic science. Hi, Charles.

0:13:09.0 Charles Rotimi: Hi, how are you?

0:13:10.2 SC: I'm good, I'm excited to talk about this. The bulk of genomic science has focused on genomes of European ancestry, we'll get to why this is a big problem in a second, but Charles, can you tell us why do we have this imbalance in the first place?

0:13:24.7 CR: I think the imbalance is the result of several factors. The initial investment in genomics was mostly in North America and Europe, and to some extent in some parts of Asia, and

also the existing cohort that genomics built upon were mostly of European ancestry individuals. So there was a legacy issue of the big cohorts and studies that were available initially were those of European ancestry and there was also the issue of scientists who have the expertise participate in genomics. Most of those scientists were initially from the European ancestry individuals, this funding, the way science has been funded in the past and also availability of the necessary infrastructure in terms of laboratories to do this kind of work.

0:14:23.6 SC: You outline in your piece several reasons that this bias is a problem, and you point to Africa as a great example of where genomic science is really missing out. Can you talk about some of the things that are missing from genomic science because we don't have a lot of genomes from African populations sequenced?

0:14:43.0 CR: We all can indeed claim Africa as our ancestral home, because that's where humans evolved some 300,000 years ago, before some of us decided to migrate to populate the rest of the world about a 100,000 years ago. Because of that long evolutionary history of humans on the continent of Africa, we have had more time to acquire more variation, so when you look at the genomes of African individuals, you see more variation there, in general. And for the most part, the variation that exists in other parts of the world are actually a subset of what you see in African population. Those individuals who migrated out of Africa about 100,000 years ago only carried a subset of the genetic variation that existed on the continent then.

0:15:43.6 CR: So there is a large proportion of what we call common genetic variations that can only be studied in African individuals or African populations. Therefore, if we do not systematically sample across Africa, we will be missing opportunity to truly understand the scope of human genetic variation in the context of human history, and also in the context of understanding biology.

0:16:16.1 SC: Genomics has a number of different diversity problems, in who does the work, the kind of research questions being asked, the source of genomes being used in research, the distribution of funding. What are some actions that are being taken to address these problems?

0:16:32.8 CR: One of the examples that are rather remarkable success story, is what we are doing under the History Africa Consortium, where that is funded by the National Institute of Health and also the Wellcome Trust in the UK. We have now successfully engaged over 500 African investigators and enrolled over 100,000 applicants, mostly from Sub-Saharan Africa, to participate in genomic studies. So that's a very successful story in terms of changing the participation of African ancestry individuals in genome science, and not just at the level of subjects or participants, but also at the level of the scientists, and we are training the next generation of scientists also, so that they can continue to be involved in the genomics studies.

0:17:30.1 CR: At the US level, there is the All of Us Program that is enrolling different ancestral backgrounds and there is major emphasis under the All Of Us Program also to increase diversity of participants in the study. So there are different initiatives, but one thing that I would want to address is really how we are moving genomics to the clinic, and that is as we get more successful, for example, there are studies going now about how we can do genes editing to treat something like sickle cell. It is important for us to make sure that we are indeed taking into consideration all of the

cultural and ethical issues that are involved in that processes, especially if we are going to deploy it into different environments in Africa or in other parts of the world, so that is critically important to be sensitive to those issues and also to develop technologies that are friendly to environments that may be resource challenged.

0:18:36.5 CR: I think that would help us to properly deploy genomics in a way that we are not going to exacerbate our health disparity or health inequalities across the world.

0:18:47.7 SC: As genomic science advances, as it enters the clinic, how is it going to benefit people who are contributing to this research from African ancestry?

0:18:57.6 CR: There are several ways that African people who are participating in genomics and Africans in general are going to benefit from genomics. For example, we already learned that a variant that is in a genome called APL-1 is very important in terms of survival of the infection due to trypanosomiasis, which we call African sleeping sickness. It turns out that that variant that is highly protective against that infectious disease is deleterious to our kidney, so right now, individuals, African-Americans, individuals with African ancestry across the world, who carry two copies of these kidney disease variants are at high risk of kidney failure.

0:19:48.5 CR: So it is important that before somebody donate their kidney to get tested for this variant, so that they know the kidney that is left would not fail. 'Cause if you accept a new kidney, you want to know the status of the donor also with respect to these variants. So that is one direct way that we are indeed understanding the impact of genomics on health of African ancestry individuals. We are finding also variants that unique to African ancestry populations in terms of diabetes and other conditions. So by participating in these various studies, Africans will indeed be represented as we deploy genetics in developing therapeutics or even treatment strategies and preventive strategies.

0:20:36.8 SC: Great. Thank you so much, Charles.

0:20:38.4 CR: Thank you.

0:20:39.1 SC: Charles Rotimi is the Director for the Center for Research on Genomics and Global Health at the National Human Genome Research Institute.

[music]

0:20:52.3 SC: We just finished talking about a lack of diversity in genomic science, in researchers, in the genome studied. But now we're going to talk about another long-standing problem in genomics, the confusion between race, genetics and ancestry. Dorothy Roberts wrote about it this week in Science. Hi, Dorothy.

0:21:14.2 Dorothy Roberts: Hi, how are you?

0:21:16.1 SC: I'm good. How come we're not celebrating 20 years of no longer looking for race in

genomes? Are people still trying to find genes for race in this day and age?

0:21:26.0 DR: Yes. There are a lot of genomic scientists who are still trying to figure out the best way to identify races genetically and there are also a lot who are looking for race-based genetic difference to explain inequities of health and some even other kinds of inequities, like in education and in violence, even.

0:21:52.2 SC: There isn't really any evidence for this happening. Why do you think people are still looking for this?

0:21:56.7 DR: Well, first of all, even though human beings are very, very similar genetically, there's only a tiny percentage of difference between human beings, there still is a lot of genetic variation in the human species. But what some scientists then do is say, "Well, we're going to look for the racial differences in that amount of variation." The problem is that all of that genetic diversity isn't grouped by race, because race isn't a biological category. Now, why do we continue to do that? I think race is just such an embedded idea in Western science, in our culture, in our society. It's useful. It was invented because it was useful for political reasons and it continues to be useful politically to explain why we have so much inequality in our society.

0:23:06.1 SC: One thing, I see a lot of is check boxes identifying race being used in biomedical studies. Can you talk about how this has been a problem?

0:23:15.2 DR: Well, the origin of checking boxes for biomedical research, including clinical trials, was originally to get more diversity among research participants. The problem is when those boxes are seen as being biological categories instead of social categories. So the boxes represent federal categories that are used in the census, they're used for collecting financial data, they're used in lots and lots of ways, and the government is very clear that they're social categories, not biological ones. But when researchers are told they have to check that box, they often confuse checking it as a social category with having to actually find biological differences between the participants who fell in those categories, and there's a lot of confusion around that.

0:24:22.7 DR: I found, when I was doing research for my book, Fatal Invention, that scientists told me that they looked for differences between races because they felt they were supposed to in order to get NIH funding. So that confusion between race as a biological versus a social category persists in science today, in this very concrete way of why, as a scientist, you need to check a box for your research participants.

0:24:57.6 SC: What are the recommendations now? What should a scientist be doing instead if they are collecting demographic data on a person and they care about ancestry?

0:25:08.2 DR: I think the most important first step is for scientists to be clear that race is not a biological category. It is purely an invented social or political category. It's not a natural division of human beings that some aspect of nature created, whether we say God created it or nature created it or evolution created it. That's all false. So if scientists could understand that it is a way of managing racialized populations for political reasons, then they can use it in the right way.

0:25:54.8 DR: To me, that's the message that I and others who are concerned about this are trying to get across, think about innovative ways in which we can measure the impact of racism on people's health and on other aspects of their lives without pretending that it's some innate biological distinction that is producing these inequalities.

0:26:20.8 SC: Can you give some examples of how studies where people are checking a box that says something about their race and then scientists are drawing conclusions from that about their biology that aren't helpful?

0:26:31.3 DR: There are a whole host of studies that have been conducted since the time of slavery that assume that black people, for example, are a biologically distinct group and categorically different from other human beings. I can give you the example of the estimated glomerular filtration rate, which is a very important indicator of kidney function, and in studies looking at kidney function in the past, researchers found that in studies, it seemed that the black participants had a higher rate of creatinine, which is important in measuring kidney function. And so in the last 20 years or so, the algorithms to figure out kidney function have automatically adjusted for black race.

0:27:29.1 DR: In other words, in the lab test, the blood test that measures for kidney function, the estimate is adjusted upward for a black patient automatically and categorically, regardless of what it was that made this finding true in the past about the particular black research subjects in that study, then there is this leap to say that all black patients should be treated automatically differently from all other human beings.

0:28:04.0 DR: We now know that that upward adjustment, which is a healthier adjustment, has resulted in denying black patients access to kidney transplant waiting lists. It's a measure that will deter them from getting specialized care because it looks as if they're healthier than any other patient of a different race. So I think that's a very stark example where the automatic categorization of people by race is harmful to patients' health.

0:28:43.6 SC: Is this something I've heard of before called race correction, where there's this built-in factor that basically says, "Oh, you're of this race, we're going to change this value for you, we're going to evaluate you differently," but of course, it was determined using faulty principles in the first place?

0:29:00.9 DR: Yes, that's exactly what it's called "race correction." And the idea is you have to adjust based on race, a measurement, because people of different races bodies function innately differently, and many of these corrections focus specifically on black patients and treat black patients differently than other patients. There's a whole slew of these measures that are corrected for race: In lung function, in hypertension, in whether or not a person should be able to have a vaginal delivery after a C-section. Many, many assumptions that are made that especially black people's bodies function differently than all other human beings' bodies. I think when I say that, it sounds... To me, it's sounds astounding that medicine would be practiced that way, but this is very commonly embedded in algorithms and diagnostic tools that are used every single day in hospitals across the country.

0:30:12.2 SC: Is this something that you feel is going to be addressed through better genomics, better genetics, is this something that we're going to research our way out of?

0:30:21.1 DR: I think we need to change the way race is treated in many, many aspects of medicine and biomedical research. Genetics is one part of it, and I think genetics is very powerful, because people believe that genetics explains a lot and that it can be the answer to a lot of health problems. So I think genetics is a particular area where we have to focus on getting rid of this false view of biological race, but it is also very dominant in other aspects of biology, in neuroscience, in cardiovascular research, in pulmonary research, in medicine. So it's across the board. And it's interesting you asked, "Can we research our way out of it?" I think that until we get rid of this idea, we will not make important advances in medicine because, to me, the reliance on this archaic view that human beings are naturally divided into four, five or six races, which is refuted by evolutionary biology, by sociology, anthropology, by law, by history...

0:31:47.6 SC: Genetics, genomics, all the stuff we're talking about, yeah.

0:31:50.6 DR: But continuing to rely on really this folklore about human creation, it is an impediment to good science. In the eGFR, for example, the reliance on race as the way that you determine kidney function has been a deterrent to finding better ways of measuring kidney function. And I think now that some hospitals are finally abolishing it. A handful have just very recently in the last year. Now, researchers are forced to figure out, well, what is a good way to do it? This is true in so many aspects of medicine, and I think the same is true for genetics. Genetics is supposed to be leading us toward personalized medicine. As long as it relies on this crude construct of race, as if it were a biological category, it is less personalized. I think that until we have the political will to stop relying on biological concepts of race, the research is not going to advance in a way that can help us end health inequities.

0:33:11.7 SC: Thank you so much, Dorothy.

0:33:13.2 DR: Sure.

0:33:14.4 SC: Dorothy Roberts is a professor in the department of Africana Studies, Department of Sociology and Law School at the University of Pennsylvania. You can find a link to her article and the rest of the vignettes in this section at sciencemag.org/podcast. Be sure to check out the rest of our human genome anniversary package. We have pieces on data sharing, things like the affordability and value of using genomics in medicine and so much more. You can find links on the episode page at sciencemag.org/podcasts.

0:33:44.9 SC: And that concludes this edition of the Science Podcast. If you have any comments or suggestions for the show, write to us at SciencePodcast@aaas.org. You can listen to the show on the Science website at sciencemag.org/podcasts. On the site, you'll find links to the research and news discussed in the episode, and of course, you can subscribe, any way you get your podcasts. This show was edited and produced by Sarah Crespi with production help from Podigy, Meagan Cantwell and Joel Goldberg. Transcripts are by Scribie. Jeffrey Cook composed the music. On

behalf of Science Magazine and its publisher, AAAS, thanks for joining us.