**The G Word Trasncript**

**Lyra Nightingale Podcast: Why Isn't Data Neutral?**

**Lyra:** I am Lyra Nightingale, ethics lead at Genomics England, and you are listening to the G word.

Through the conversations we have on this podcast, we hope to bring the benefits of genomic medicine to everyone. Today, we will be discussing the ethical, legal, and social issues of diversifying data. We're joined by Professor Anneke Lucassen and Dr Faranak Hardcastle of the Clinical Ethics Law and Society Research Group. Anneke and Faranak are two of the authors on a literature review commissioned by Genomics England to look at the ethical, legal, and social issues in diversifying data. The review has been instrumental in helping the team think about how we design and operationalize our diverse data program, which has the ambitious aim of sequencing up to 25,000 underrepresented participants over the coming years.

So, I am joined today by Anneke and Faranak would like to both introduce yourselves, please.

**Faranak:** I am a research fellow at the Clinical Ethics Law and Society Group that is led by Anneke at Oxford, and I am interested in exploring how technologies and societies shape each other and evolve together, and how we can intervene in this evolution to direct it towards a point where their benefits are equally distributed.

**Lyra:** And Anneke you are no stranger to Genomics England . Do you want to tell us a little bit about your background? And also some of the ways you've been involved in genomics England?

**Anneke:** Great, thank you. Yeah, my name is Anneke Lucassen, and my background is that I trained as a physician, and then I did some research in the genetics laboratory discovering genes, and I wanted to translate that into clinical practice. And I found that in doing so, the most interesting issues were the ethical, legal, and social issues that really needed to be addressed for that translation and implementation.

So, I'm a professor of genomic medicine at Oxford now after moving there recently, and I am Director of the Centre for Personalised Medicine, and I run a research group, as Faranak mentioned, looking at the social, ethical, legal issues that we need to think about when integrating massive developments in technology and science into practice and society. So it's a really exciting interdisciplinary space to inhabit, and this review was one of the really great ways to exercise that.

No stranger to Genomics England indeed, I was involved with Mike Parker in the very early discussions before Genomics England was formed. And I have been on its Ethics Advisory Committee ever since, being involved in science-wise collaborations that Genomics England have done about public consultations. And I'm, at the moment, involved in the newborn screening ethics subgroup.

**Lyra:** So, you're very familiar with some of the issues we come across at Genomics England, and I suppose my involvement with you both started when Anneke. I obviously met you through our Ethics Advisory Board, but another way we've worked together is that I am the Ethics Lead for our Diverse Data programme. So, the vision of our Diverse Data programme is to make sure that all patients and participants, regardless of their background, receive the same quality of genomic medicine. And we know that at the moment, that isn't the case; there's a lot of bias in our datasets. People from a white European ancestry are massively overrepresented. And so the insights that we get from genomic medicine don't work equally for everyone. One of the aims of the Diverse Data programme is to sequence up to 25,000 participants of non-European ancestry to remedy some of that imbalance.

Now, when I was faced with this, when I first joined the programme, I was kind of slightly overwhelmed with fear of all the ways we could actually make things worse with very, very good intentions. So, for example, if some of my fears were that we might end up emphasising diversity of ethnicity over and above socio-economic factors or deprivation. I also really worried about this idea of emphasising difference between groups that could potentially give, say, a biological underpinning to this very social idea of race. One of the ways I sought to address these fears was to ask for a literature review, looking at all the ways that programmes like this in the past, or the issues they've come across in terms of their legal implications, social ethical implications are kind of the unintended consequences of programmes such as this.

**Anneke:** Lyra, I think that is so important to ask the question about how we get truly diverse data because it's going to be more complicated than just going out and sampling people. So, we are really pleased that you asked that question, and I think one of the thorny issues there is how we use, or how we think of terms like ethnicity, race, and ancestry. We're very used in the UK to answering boxes about our ethnicity in a very prescribed tick box format. But the issue is, how do we use that information? Do we use it solely to infer genetic ancestry? Or do we appreciate it as something a bit more complex than that, that also involves socio-economic and sociodemographic information that we don't ask specifically about?

So, I think until we appreciate that those are really interwoven and that whilst we have a tendency, sometimes I think, to ask ethnicity data and then only interpret genetic information from that, we're going to run into difficulty. So that's why we thought it was so helpful to be able to do this review to look a bit more at the complexities behind that. But I'm going to hand over to Faranak to give us a bit more information about the review.

**Lyra:** Yeah, I'm really glad you raised that Anneke, because I'm going to focus a lot of this discussion around what these difficulties are, and more importantly, what we can do about them at Diverse Data, which I might refer to as DD at some point when talking about diverse data, refers to that. Faranak, if you could tell us a little bit about how you even approach this question, that would be really interesting. Thank you.

**Faranak:** Yes. So, we were an interdisciplinary team based at the Clinical Ethics, Law, and Society Research Group across the Universities of Oxford and Southampton and the review had three main components. It had a systematic element, a consultation piece, and a narrative review. And I can say a bit more about what each element involved if you like.

So, in a systematic element, we searched for empirical studies with a primary or secondary focus on diversity in genomic data, or the inclusion of underserved groups in genomic or health studies, and the primary focus or secondary focus on ethical, legal, and social issues related to data diversification.

We then held a workshop with some of the experts in the field and tried to identify some of the gaps in the literature that we had looked at and then supplemented that with a narrative review. And I just want to add that the search that we did for the systematic part of the review had some limitations.

One of the limitations was that most of the papers that we reviewed were from North America. Also, another limitation was that our search mainly focused on under-representation that was based on gender, race, and ethnicity. That leaves out other underserved groups, such as children, the elderly, psychiatric patients, prisoners, and so on. This links to the issue in attempts to diversify, which is the categories that we tend to use for participant recruitment may not accurately map to ancestral categories that are needed for genomic analysis that Anneke just mentioned earlier.

**Lyra:** What are we talking about when we are talking about ancestry, for example, and this categorization of ancestry or even ethnicity or race, and how do they all fit together?

**Anneke:** There are no tight definitions universally accepted, of those terms and that is one of the issues. Race and ethnicity are both seen as social constructs that are more than your biological or genetic background. And the problem, as I mentioned earlier, is often that that bit is forgotten. We assume that if we tick the ethnicity box, or ethnicity (which I think is more commonly used in the UK) and race (perhaps more commonly to mean something very similar in the US, which is relevant to our review, I think, as well). But I think for here, we should use race and ethnicity as interchangeable.

And then, genetic ancestry, I would say, is different in that it is looking at genetic variants and how they differ according to recent migration across the globe. When I say recent, given that we all started off, or humans started off, in Central Africa and then we have migrated across the globe over the last 10 years. What we're looking at is genetic variation that represents the sum total of that; that's happened differently in different geographical regions.

So, I think from the genomics point of view, that's a really key question, isn't it? How is normal genomic variation different depending on your ancestry? And that isn't answered by the colour of your skin; that isn't answered by what you perceive your ethnicity to be. It might have some vague correlation, but it's much vaguer than most people expect.

So, when we're answering questions about whether we're white British, or traveller, or from Asia (I can't remember the standard questions on those ethnicity questionnaires), they're very confusing questions to answer, because your ethnicity is what you self-identify as, and it's cultural, as well as geographical, as well as your religion, your beliefs. Those sorts of aspects aren't measured by those categories.

And so, it's a much richer, vaguer term than the one we really want, which is: what is normal genetic variation for this particular person? So I think there's huge problems with trying to conflate the two, and I know there's other work around some genetic workarounds, but they still don't really get at the fact that your genetic ancestry is one question. And then your socio-cultural identity may overlap a little bit with that, but it's quite separate.

And if we're talking about diversity, we want more than diversity of genomic information. We want diversity of communities and groups, etc.

**Lyra:** I think, in genetic ancestry, it is genetic similarity to a reference genome. And so, although ancestry is a valid research tool for exploring how genetic variants are distributed across populations, the problem is when we communicate this to our racialized worlds, it kind of gets lost in translation. And we then assume we are talking about race and ethnicity. And that seems to be one of the problems as well.

**Lyra:** Thanks, Anneke. So, what you have identified is that we really do need more diversity later. And if we don't have it, the consequences are absolutely disastrous in terms of putting people through unnecessary diagnostic tests, for example, or insights not working, and pharmacogenomics being less effective. But one of the things that you've pointed out is that having more diverse data in and of itself alone is not ethical. And actually, there's a lot more that needs to happen than just diverse sequencing.

One of the things that came up in the review is that if we just carry on with this approach to sequencing, using existing research practices, we might actually exacerbate some of the existing inequalities. Faranak I don't know if you'd like to speak to that a little bit.

**Faranak:** Yes, that is right. There has been a recognition that the divides are getting bigger, and we need to address health inequities. And one of the reasons behind the inequities, as Anneke just extensively explained, is that the data we use to advance science and technologies are not diverse. New developments may only benefit those who are already privileged. So, it has been more than a decade now that scientists and clinicians are calling for more diversity in genomic data.

In response, there has been a great deal of enthusiasm about diversifying genomic data, which is great, but it can get a bit confusing when well-intentioned people equate diversifying data with addressing health inequities. Because addressing health inequities needs more than just having diverse data. And getting those datasets more representative needs to be a part of making the whole enterprise of genomics more diverse, not the goal in itself.

The key message from our review is that diverse datasets shouldn't be an endpoint in themselves just collecting genomic data from people with a range of ancestries doesn't address the diversity problem. We also need to think about who is conducting the research; are all stages of the research processes conducted with those people from the underserved groups? What tools, methods, and categories are being used? What knowledge is being produced there, for example, what information was counted as scientific and what has been discarded? Are we also looking towards ensuring that the benefits of that knowledge can be fed back to the diverse populations and so on?

And so, one of the things that we discussed in the report was that the call for diversity shouldn't be just about the data. It should also be about the type of knowledge that is being made, tools and methods that are being used, the workforce and infrastructure that is in place, and also the disciplines that are getting engaged in the research. And these are all very much connected to each other.

So, the lack of diverse data is one of those things that you can't really tackle from one angle and in silos; it has to be all thought through in terms of the bigger picture. And if we ignore this bigger picture in our research practices, we might end up worsening some of the issues for some people.

**Anneke:** Very succinct example of what Faranak said: that people have been calling for more diversity for a long time and there are massive efforts underway around the globe to increase the diversity of our genomic understanding. But actually, if you look at a site called gwasdiversitymonitor.com, it's a very nice graphic site of all the GWAS (genome-wide association study) studies. So there are genome-wide association studies done globally and to date and that shows the polarization getting worse.

So more than 95%, I think it is, of all GWAS studies have been done on people of European, probably more Northern European, ancestry and despite those calls, that's actually got worse rather than better. So that, I think, to me really illustrates starkly how we entrench our existing position, unless we take a real step back and have a look at, as Faranak said, in the round, not just about getting another sequence, but thinking about the whole process from start to finish. We'll start and ongoing, and then incorporate that in our attempts to diversify the data.

**Lyra:** I'd really like to come on to that process that you speak of more, especially in terms of co-production of knowledge. But one of the things that you touched on, Faranak, is the kind of questions we ask and who's asking the questions and Anneke, I think you mentioned as well, the data that we're collecting, the calls for the data that we're collecting. And one of the things that came out from the paper for me was this idea that data isn't neutral.

So, I think there's this real, I don't know the right word here, but an assumption that science is very neutral. It is not involved in politics. Same with data, so that data can represent the entire world. It is factual and I think your review highlights that that's actually not the case. So, just collecting diverse data alone, we need to be careful about the assumptions we're drawing from that data and Anneke, I think you touched on it in terms of what data are we not collecting.

So, I don't know if either of you are willing to share a little bit more about this kind of paradigm that data is neutral and represents facts, whether you agree with that or disagree.

**Faranak:** Yes, that's right, Lyra. Data are not neutral or objective, because they do not exist in a vacuum and they are produced by processes that we put them through. We curate them, we select them, collect them, clean them, edit them, and classify them, and then end up analysing them. We live in societies that are not neutral and are assembled by unequal relations themselves. So our social, cultural, political values, and assumptions that we make about the world all the time all get embedded in this data. A quote that I like with regards to this is, "data are not raw; they are cooked according to a particular recipe."

This is not to say that data are useless, and we should not be using them. It is just to emphasize that they need to be understood within that context. They can be misleading if they are taken out of their social context. So, for example, during the COVID pandemic, we saw that black people were disproportionately affected by COVID and some researchers used COVID data to conclude that there is a genetic risk factor in black people that makes them more susceptible to COVID. We cannot ignore the fact that being exposed to structural and everyday racism can have negative health outcomes. So, what is perceived as genetic predisposition, in fact, is social disadvantage. So we need to be aware of that social context to be able to be both ethical and scientific. I mean, I can go on. Anneke, would you like to jump in?

**Anneke:** The only thing I would say is that you might have a particular data point that people would argue is subjective. And scientists may then say, "This is all I can do. I cannot add in all the, you know, the sort of vague things. I'm sticking to my nice scientific approach." But as Faranak says, if you have multiple data points, it depends on how they're cooked.

So, I think there is a bit of a conflict between a particular data point maybe in itself being as objective as it's possible to be or as neutral as it's possible to be. In the whole, we need to pay much more attention to how they are collected.

**Lyra:** It is so interesting, it's something I'm thinking about a lot, as well as, what conclusions are we going to draw from the data that we collect? And do we have the right data or the right assumptions to be able to even draw some of those conclusions? So, Faranak, you gave that really classic example of COVID which is such a deeper question: purely ethnicity versus outcome?

**Anneke:** And even purely ethnicity is really difficult, isn't it? Yes, people self-declare their ethnicity. And yet, we then exclude them from studies on the basis of that self-declaration. I am not arguing against self-declaration, but I am just saying that, you know, we take that as a very objective fact, but it depends on how I feel. I fill in my own questionnaires differently, depending on what I feel.

**Lyra:** Yes, what you feel the questions after, the level of trust you might have in the person asking the question, there's a whole host of issues there.

**Anneke:** I was just going to say, you know looking at examples of large established biobanks, their research studies have excluded data based on a sort of very crude classification of genetic ancestry that, in fact, is already different. The questions about ethnicity are now asked in a different way than they were in the past. That just gives you an idea of the noise in this system and what you can start excluding or including based on questions that you're not currently asking. That's one, to me, very good example of how data can't be neutral, because it depends on how you're filtering them.

**Lyra:** Yes, absolutely and one of the things we've been doing here at Genomics England is trying to look at and identify the bias within our own pipelines. So, for example, if someone comes in and they have a specific character in their name, does that data get cleaned out, for example? There are so many different elements to this before we even get the data in our research environment.

Anneke, a challenge to you would be: then, we're diversifying our datasets, we have all these REC approvals already in place (Research Ethics Committee), you're really looking at the ethics of what we're doing. We're surrounded by laws to kind of protect the sensitive nature of health data or ancestry data. There is plenty of regulation around. Why is that not enough? I know that your paper pulled out issues that actually these regulatory practices can sometimes obstruct ethics, which sounds contradictory. So, could you tell us a little bit more about how that plays out?

**Anneke:** Yes, it is important to recognize that this regulation and legal rulings are really important, and an important reaction to past injustices. So, you know, the whole idea of a Research Ethics Committee is very much post-Second World War and the atrocities we saw there. In the clinical sphere, there might be all sorts of regulations post scandals that have happened – Shipman, for example, or organ retention, those sorts of things. So, it is really important to acknowledge that they're there for a reason, and they're there for a good reason. I think what sometimes happens is that they're asked to do all the work.

And what we have been working on in our group, in fact, in collaboration with Professor Bobbie Farsides' group in Brighton, is on a concept of ethical preparedness, to think about what is necessary beyond rules and regulations. What I think sometimes happens is that people think, "Well, there are laws here and regulations. So that covers everything, as long as I follow, take that law off the shelf, and make sure I've ticked that box, then everything else will follow.

**Lyra:** Yes, as long as we can get it through REC.

**Anneke:** Yes, and RECs are often right at the start of a study, and they don't have much to say about what happens when the study is underway – no annual reports or whatever – but they don't really help with looking at ethical issues in the study itself, or after a study has finished. They are more and more, I would say, a form of governance, an important form of governance; they don't really look at contextual ethical issues that arise. So, ethical preparedness is really trying to say, how can we better include the nuances of particular situations – be that research or clinical practice – and empower practitioners, researchers, and the public to feel that they themselves can examine and judge what are the particular ethical issues in that situation that aren't covered by off-the-shelf solutions? And how can they be empowered.

And I'd say, that might be very different in a clinical setting than in a research setting. But of course, what Genomics England highlights so beautifully is that those two areas cannot be neatly separated any longer in many modern-day endeavours. That whilst on the one hand, genomics is a very exciting research endeavour, it interacts very closely with clinical practice now, in a way that really means that research and clinical practice are alloyed together. That can't be covered by Research Ethics Committees, because that's for research only, and it can't be covered by clinical practice because that's for clinical practice only. So, we really need new approaches to be able to look at this alloyed together, hybrid area, I think, and so it's really interesting to think how ethical preparedness in that setting might be able to help. It may be off-piste a little bit, but I think it's very relevant to the review on diverse data as well.

**Lyra:** No, I do not think you have gone off-piste. I think one of the examples for me that really highlights everything you've just said in terms of the protections for the individual and their autonomy versus the risk of unethical practice could be in the way that we do our consent documentation, for example. So, in England, at least, there is an awful lot of emphasis on consenting the individual, looking at their autonomy and their choices. That doesn't always take, or doesn't very often, if ever, take into account the potential for community harm or this idea of community consent to the acceptability of research.

**Anneke:** I think that is absolutely right. I think consent is often seen as another tick box, isn't it? It's sort of it's used as a passive verb quite often, you know, has the patient or the participant been consented rather than have we had a consent conversation that's covered everything that they need to know? Worked together in a system of ethics is more than just about, you know, tick box in that patient's autonomy, as you say, about how it sits within communities. Genetics very definitely, often goes beyond the individual in terms of familial implications. In my clinical practice, by often, you know, a consent conversation with lots of tick boxes, in small font, going over six pages, will never cover all the possible options for that person and their family members

**Lyra:** I think what is important to different people or different groups, if we're looking at grouping people, is going to vary massively. I think one of the huge emphasis of your view is on co-creation and co-production to understand more about that, but we will come on to that. It's a big topic, and I really want to hear more about it. One of the things that the review suggested that I found really interesting was this idea that researchers need to develop more cultural humility. To me, that speaks slightly of understanding the ground that you're walking on in the first place in terms of the historical context that we've talked about. It seems cultural humility is more than that, and Faranak, I don’t know if you want to talk on that point. What is cultural humility in research? What are we talking about there?

**Faranak:** Sure, yes. So, the review highlighted how important it is to acknowledge and incorporate the experiences and expertise of potential participants, especially those from underserved groups. This, as you say, Lyra, requires understanding cultural values and being reflexive. Cultural humility is used often to emphasize the importance of reflexivity, active listening, and taking responsibility for interactions on the side of researchers and research institutions. That's cultural humility, but as you say, we could potentially worsen things for some people if we practice research that is exclusionary and doesn't consider the cultural settings in which potential participants are situated.

**Lyra**: I think a good example of that would be research where we just go in, take what we need, we've diversified our data set, the end. And I think that your review pulls that out as an example of a colonial research practice, a very traditional extrapolative model, where research is potentially in a more powerful or more dominant position, come in, take what they need. The community never hears from them again. So, I think this is a good segue into my next point, which is there are many harms. How can we mitigate against them? What is the way forward

**Faranak:** This is a difficult question. I think we found some nice best practices from other countries that have been trying to co-produce genomic knowledge. So we talked earlier about how that's not enough to have diverse data on its own, and we also need to bring in diverse views on what counts as knowledge. So while co-production, I guess, it's the idea that we research with participants as opposed to on them. That's what we meant in our review. I think that doing co-production, in a true sense, is really difficult, and I think what we need to ensure is that potential participants are seen as active researchers and knowledge producers. If we don't go with such a mindset, then it becomes really easy for our co-production or engagement activity to become tokenistic. That in itself can risk exacerbating existing problems. We found some nice practices in our review that had been trying to co-produce genomic knowledge, but in the context of the UK, it's maybe that we need to really try and work out what works best for a super diverse society like the UK. Again, because best practices also talk about, you know, going to a specific community and just trying to get them engaged in research. But how is it started? From the beginning, in a very diverse society? How can we co-produce naturally? And how can that naturally occur? Is it something that we have not really explored yet? I think there's a lot that we can do in that area. And part of it perhaps is to, as you talked about the difficulties of what is meant by ethnicity, race, diversity, community, all of these are questions that we can try and answer in discussion and in conversation with all the people that we want to have there later.

**Lyra:** So itsounds like co production has a lot of overlap with engagement, but it isn't the same as just purely engaging with groups about their ideas, right. It's something over and above that, Anneke I can see you shaking your head. I don't know if you want to say a bit more.

**Anneke:** Well, I was just thinking, in terms of you asking what's the way forward, I think the way forward is firstly a greater awareness. I would say that I've been interested in these issues for a long time, and yet I still learn new things from the review. So I suspect that other people have room to improve their awareness as well. So it's not just, 'oh, we need to sort of think about co-production and just think about everything in the whole.' I think what's really important there is that true interdisciplinarity is really important here. We talk a lot about interdisciplinary research and interdisciplinary approaches, but we haven't yet got the whole way. I think much more focus on really integrating clinical practice, scientific research, social sciences, philosophy, and lots of different approaches to this question and truly interweaving those all together. Think about what they have to say about this is going to be really important. What we don't want is just another set of rules. We don't want, 'oh, this study worked really well in this particular research. So let's apply that across the board so that everybody fills in a tick box, or let's create a new law that says you must do X, Y, and Z.' It's got to be something broader than that. I'm not suggesting that laws are not necessary, they may well be, but I think we need something much broader and richer to cultivate an awareness. We will do that by raising these issues, but also by really trying to inhabit that interdisciplinary space in a more effective way than we've been able to sofar.

**Lyra:** Yes, thank you. And is that similar to the emphasis you've talked about on this idea of diverse knowledge that sounds very similar.

**Anneke:** That's probably right isn't it Faranak?

**Faranak:** Yes, it is very much on those lines. So it was just to say that we need to bring in diverse views on what counts as knowledge. Many of our conventional tools and methods are exclusive. We can only have true diversity if we try to make them inclusive and diverse as well. So the enterprise of genomics needs to be more diverse as a whole, is just to emphasize that again.

**Lyra:** That even came up in your review, right, the way you conducted the review, when you were looking at the papers that were predominantly written in English. Correct me if I'm wrong, but part of the reason you brought in the focus group was to surface some of the issues that you can't get through these traditional methods alone.

**Faranak:** Yes, it was a workshop it was exactly for that reason, because the literature didn't really cover much. A lot of the synthesis from our findings comes from the consultations we had with the experts on the narrative review that we did afterwards.

**Lyra:** Thank you. So, we've talked about co production, could you give me some more examples of positive action that we at the diversity programme could take that was surfaced by your review in mitigating some of these harms?

**Faranak**: I think the idea of awareness that Annika just touched upon is really important. That awareness, I think, needs to be across everything, all stages of the process. How, as you said earlier, Lyra, that data are not themselves neutral, but also, our tools and methods are also not neutral. They are also political. For example, classification systems that we just talked about, they are often sites of a political struggle. A lot of times, we don't know that in the beginning because our tools and methods are kind of presented as very technical, as that veneer of technicality that almost feels like as if we can't really get through to what's underneath. When they are presented as such, it's almost implied that they are neutral and objective. The problem can be that there could have been political agendas just presented as something technical, and as we use these systems, and they become part of the infrastructure, and they become normalized, then it's really difficult to go unpack all that political stuff that's underneath. So it's that idea of that Annika mentioned about raising awareness, I think is so important. We need to occasionally just have a pause and open the bonnet, have a discussion about what's inside, and then take the next steps. That is really important.

**Lyra:** Yeah, I love that idea of pausing and taking a look because I think one of the things that really came up for me from the review is that, in my experience, there's a focus on this so engagement or ethics at the start, and what you're talking about there finds slightly, is this idea of a feedback loop. Is this even working? Are we doing this right? I think your review really focused on the idea that ethics isn't just at the beginning, we need to co-produce and look at our ethical preparedness or our awareness the whole way through, even when we're talking about publication of research at the end of your process, feedback loops, capacity building.

**Anneke:** I was just thinking of our next analogy that she attributed briefly earlier about data being baked in particular ways, and things that we might want to suggest is that we check the oven. We take the cake out the oven halfway through, and if it's not actually risen properly, or it does not look very tasty, then we start again. And that does not mean never baking again.

**Lyra:** Like certain people might be like that cake looks great.

**Anneke:** Exactly, so it may be keeping the same ingredients but having a discussion about how if they're put together in a slightly different sequence, you might come up with a different product or that if it doesn't rise properly, you might start again and just change the conditions a little bit. I don't know if that analogy works, but I like to think of being able to do that. Not saying, 'Okay, well, you're so depressing about diversity, we're never going to be able to do that.' So, what you're saying really is the opposite. We're absolutely not saying that. We're saying that thinking about these issues will help well-established databases as well to make sure that these political, sociocultural factors that Faranak so nicely articulated are incorporated better than they are so far, so that we don't get this sort of ever-increasing skewing of diversity despite getting more sequences from people of different ancestries

**Lyra**: I like what you said there about it's not always starting over or new ingredients. We have some of these tools, like we're really familiar with governance at Genomics England, but how can we adapt that to meaningfully Power Share? It's not that, I mean, I'm sure there are some cases where we need entirely new and novel solutions that we will learn through co-production. But often we have these tools already. We are just applying them in a way that doesn't work for everyone. I have a slightly more personal question. Annika, you mentioned earlier that you've been working in this space for a while, but the review still surfaces things that surprised you. I would love to know some of the things that, as a seasoned kind of ethicist looking at these issues a lot, came out for you that was novel or that you might not have had at the front of your mind.

**Anneke:** I think the most surprising thing for me was, despite all the major attempts that have been going on around the world to improve diversity, and you know, there's lots of really exciting new biobanks being set up, etcetera, etcetera, that the skewing of data towards Europeans was getting worse rather than better. I had not appreciated that. I think that comes from using old resources over and over again. It's a bit like a systematic review, I suppose. If they're all of the same data, then you might just get the same message in slightly different ways. I think what was really interesting was the review was a great way of bringing real interdisciplinarity. So, Faranak, who is really well-versed in science and technology studies, that's your background, isn't it, and lots of social scientists were involved. This was also really looking at the technical stuff and how they implemented things and the nickel sides. It felt like it was a really good way to sort of interdigitate things that were known in particular silos but we've got the whole together.

**Lyra:** Yeah, and I think that's one of the things that I loved about it as well. For me, although some of the findings were things that might seem more obvious to those working in the space of ethics or social justice, it's a really great evidence base to convince people who might see these issues as more neutral to science and data as more neutral. So, thank you so much for that. You've made my job a lot easier.

**Anneke:** Whatit also highlighted was that the review in itself, as Faranak already said, was skewed. Most of the papers were from North America. That's a really important point that we need to remember going forward.

**Lyra:** So, Faranak, your review raises a lot of the issues that are more structural or historical, including issues such as structural racism. Now, one of the arguments I've heard before is, 'We're not here to solve structural racism, we have quite a small sphere of influence. That is kind of not our problem.' Do you have a response to that, or things surfaced in the review that you think could be helpful in that?

**Faranak:** Yes, thanks for asking that. I think that a lot of people might say that we can't solve these issues overnight. What we can do is just to make things better by innovation and making new technologies that just improve things slightly. The point here is not to reject the benefits of science and technology, as Annika previously said, but just to say that our research practices need to be always contextualized. If we close our eyes to this context or say that it's not relevant to us because what we're doing is apolitical, then we run the risk of worsening the issues for some people.

**Lyra:** So if we were ignoring structural racism, and operating in a vacuum, we're actually going to exacerbate the issues that we're trying to fix that's really important.

**Anneke:** As scientists working on genome sequencing, it's very easy to think, 'I can't be involved in that. I cannot be responsible for that. As Faranak highlights so well, actually, you can. You can play a part in that conversation.

**Lyra:** You can play a part in who's involved in the conversation, which voices you're listening to. Thank you so much.

So, I have grilled you both on your paper. I recommend everyone reads it, and we will put a link to it. Can I ask you to finish this podcast by either sharing your favourite thing from the review or a top tip for me doing my job in the diversity program or for Genomics England generally? I would love to hear your kind of one takeaway for people working in this area.

**Anneke:** My takeaway message would be, we need to work on awareness of these issues. As Faranak says, so that nobody who's involved in this sphere poses areas that are thinking of not relevant to them, that we help people have conversations, that we think about how interdisciplinary approaches can really shed some new light and keep the conversations going.

**Lyra:** Any other top tips for the DD programme? I might put words into your mouth here and say, having heard you speak about these issues? It seems like co production might be a really key message for you.

**Faranak:** Yeah, I think co-production is absolutely important in all stages. There are so many examples, like the one I heard last time from a patient, saying that they were invited to a patient group, and they were trying to communicate some of their problems and thoughts. However, the people who were running the engagement were saying, 'Oh, actually, that's not relevant to the questions we're trying to ask. We want to ask X, Y, and Zed.' It feels like the mindset needs to be that we consider all of the participants and patients as active researchers and knowledge producers, and really take everything they say into consideration. We should bring them on board to do research with them, as opposed to on them. So, yeah, it would be really fantastic if that could happen.

**Lyra:** Amazing. Thank you so much. Annika as someone who sits on our ethics Advisory Committee, I trust that you will hold us to account on as many of these issues as you can, and I will be doing the same. Thank you both so much for your time, and I look forward to working with you again.

**Anneke:** Thanks, Lyra. I really enjoyed the chat.

**Faranak:** Thanks for inviting us.

**Lyra:** Thank you very much for tuning into this episode. If you've enjoyed listening, giving us a five-star review really helps others find out about the podcast. And if you have any suggestions of topics or guests do get in touch with us at podcast@genomicsengland.co.uk.

Join in on the national conversation of genomics. See you at the next episode of the G Word.