Genomics 101: What is the National Genomic Research Library transcript

**Naimah:** Have you heard of the National Genomic Research Library, and do you know what it's used for? Today I'm joined by Natalie Banner, who is the Director of Ethics here at Genomics England, and she's going to explain some more.

So, first of all, Natalie, can you tell me exactly what is the National Genomic Research Library?

**Natalie:** The library is a collection of data that we hold about the whole genome sequences of our participants. It's all held in one secure place for researchers to access. Every participant has consented for their data to be there, and it's a really phenomenal resource for researchers who are looking for the causes of cancer and rare conditions, and they're hopefully trying to find new ways of diagnosing and treating those conditions.

**Naimah:** Can you delve a bit deeper into what it's used for and what are the benefits of it?

**Natalie:** Sure thing. Well, we have a number of projects, and bear in mind, this is academic researchers as well as companies and commercial researchers. They're all looking at data that's in the library to find new treatments, potentially new drug targets, for example, for cancers and for rare conditions.

So, a good example of this is a recent study by researchers in Cambridge. They looked at 12,000 NHS cancer patients and they'd had a whole genome sequence. They did a DNA analysis of thousands of tumours from those patients, and the researchers found lots of clues about the potential causes of those cancers. They detected patterns across lots and lots of individuals, and it was really important, they were looking across a really wide range of people. They found what's called signatures. They provide clues about whether that patient might have had some exposure to environmental causes of cancer, or they had some cellular issues going on that might cause their DNA to malfunction and then subsequently cause a cancer.

So, what this team did is they spotted 58 new signatures through this research. And that suggests there are some additional causes of cancer we don't yet fully understand. So, this is basic research that can identify some flags. They can set the scene for where the research then needs to develop. For example, that could be developing new drugs that are specifically targeted to particular variants and potentially new treatments.

**Naimah:** The data that you mentioned within the library, where does this come from?

**Natalie:** Well, this is all data that has been collected often from patients who are undergoing a clinical test in the NHS. For example, if you get your whole genome sequence test, if you have cancer or some types of cancer or a rare condition, you'll be offered the opportunity to participate in this research and to allow your data to be included in our library, the National Genomic Research Library.

We also have data from our fantastic 100,000 Genomes Project participants. They signed up to the study over the past few years just as Genomics England was being set up. They were really foundational to the whole enterprise of Genomics England, and they contributed their data to allow it to be used for research purposes.

So, all of this data has come from the sequencing of someone's entire genome. And then at the same time, we also collect their ongoing health records as well. Now, those ongoing health records allow researchers to be able to look at the potential connection between the genome that they've sequenced and the variants they have in their genome and their ongoing health as it develops over time. So, you can kind of compare and spot patterns between symptoms, health conditions, and the genetic variants different individuals will have. This could be their ongoing health records. It could be data, including samples, tissue biopsies, blood tests, those sorts of things, and all of which, when gathered together, enable researchers to sort of find and spot patterns that could give us some clues about the causes of disease, to see connections between genomic variants and the phenotypes or the symptoms that individuals experience.

**Naimah:** Can anybody access this data?

**Natalie:** Well, the people who can access the data in the library, and this is both academic and industry partners. We only allow applications from those that we at Genomics England have a relationship with. So, for the academic community, these researchers are signed up to our - what's called our GECIP (Genomics England Clinical Interpretation Community). And for the industry partners, we have what we call our Discovery Forum. So, this is a collection of partners from the pharmaceutical industry, from biotech, and if we think they're working on areas that could help find new treatments and diagnoses, we will agree to work with them, and they are then allowed to apply to access the data that we hold. So that's the first stage. We only allow applications to access data from people who we already have that relationship with.

Then we have an independent Access Review Committee. Now, this looks at applications from industry and from academics, and they assess whether or not they think that application and the proposal from these partners is going to be the right sort of thing. That access review committee does contain members of our participant panel. So, it's really important to us that our participants, those whose data we are talking about here, are actually a really integral part of the application process. And they look at, well, is this use of data going to be safe, secure, responsible? Are they looking at questions we think are going to really help potentially benefit participants down the line? And that could be early-stage research or it could be something that's quite close to being applicable in the clinic.

**Naimah:** So, these people that you mentioned, what do they do with the data whenever they can access it?

**Natalie:** Well, the main thing to think about when you consider the vast amount of data in the National Genomic Research Library, we're talking absolutely huge amounts of data, is the interest that researchers have is in patterns. They're not really necessarily looking at individuals. They don't need to know who individuals are, for example, but they're looking to see across a really large number of people.

In genetics and genomics, you're very much looking for a needle in a haystack. Here you're looking for tiny signals and the variants and relationship between a genetic variant and potential symptom that could indicate something that could be helpful for them to sort of subsequently target new drugs and potentially new treatments. So those researchers are looking for patterns in the data. It's very huge and very complex. So, they'll be looking for these connections between variants and potential symptoms in all sorts of different ways. Lots of different techniques, very sophisticated analytical techniques and statistical techniques, to find these very small connections and patterns.

But the key thing is the purpose of looking at this data, it has to align with our acceptable uses policy, and that's something that sits within the research regulatory approval that we have from the Health Research Authority, and the governance I mentioned around our access review committee and all of that is to ensure that only those who are doing research that we think could have that potential benefit for diagnosis and treatment are allowed to access the data.

**Naimah:** And all of the data within the National Genomic Research Library, is this useful?

**Natalie:** It's hugely useful, and actually one of the great things about having all of this data in one place is that, for example, if an individual's participating in the library because they have a particular condition, say a kind of cancer, we can actually enable that data to be used to look for other things as well.

So, to explain, a really good case in point here was the fantastic research study with the Genomic Consortium during the COVID pandemic. This study looked at the genomes of over 7,000 patients in intensive care units in the UK during the pandemic and they compared it to people who'd either not got COVID or who had a mild form. Actually, they used a lot of data from our hundred thousand genomes participants. So, they hadn't signed up to the library because of COVID. They were part of the library for something else. But when the researchers looked at and compared data from those participants with those who had severe covid, they could actually find 60 new genetic variants associated with severe COVID-19, including some that related to blood clotting, immune response information. So, that meant that it was all fantastically helpful in discovering potential treatments for covid and other respiratory diseases.

So, the important point here was that the participants that were involved had actually originally joined the library for other reasons. They had a different condition, but they were very, very useful to be able to help us find something very insightful about COVID.

So, this data is incredibly useful for a wide number of different research areas beyond the specific reason why people signed up in the first place. That's what makes it such a phenomenal resource as a library.

**Naimah:** Can data be taken out of the library?

**Natalie:** The way we work in Genomics England is we have a secure data environment, and what that means is you can access the data, you can come to the data as an approved researcher, but you can't take it away with you. You can't download it. It can't be sort of sent on to third parties, for example.

What we have is we call it an airlock, which means that you can only take data out of the library once you've done your analysis within that secure environment. So, there could be reasons you want to take your analysis out. Let's say that you've done some research and you need to publish because you've discovered something new and exciting and you want to be able to share that information with the wider community. You can only take data out that's been approved by our Airlock Committee. They basically check to make sure the data that's being extracted isn't about individuals. It's not individual level, so it has to be sort of summaries, aggregate statistics, figures, those sorts of things for publication purposes.

Now, there are some exceptions to this rule where there is an appropriate consent in place. For example, we support clinical collaborations between clinicians and researchers, and they may want to publish something new and important, but it needs quite detailed information about a participant, quite granular information in order to be able to sort of publish and show the novelty of the research they've done. In those clinical collaboration cases, they will explicitly ask for the patient's consent to include that granular detail. And it's the clinician treating that patient, so they have that relationship with them. They'll ask if they can publish that detail in an article. And in those circumstances, because the participant has explicitly said yes, we can allow that detailed data to go outside of our research environment.

In some other cases, our participants may have signed up to another study as well as participating in the National Genomic Research Library. For example, we have a few hundred participants who are also members of UK Biobank, and they've signed up to their terms as well as part of an ongoing data resource. We say that these participants have a dual consent, and in those cases, it can be really helpful for us to allow the data that we hold about those participants to actually be shared with that other study. And then they can do richer analysis with other data that they hold. So, in the case of Biobank, they have additional data on those participants, and they want to be able to sort of analyse all the data together. In those circumstances, we can send them the data on those participants. So, what's important to note is in all those cases, everything that's permitted is entirely determined by the consent the participant has provided.

**Naimah:** So, what if a researcher finds a potential new diagnosis in the National Genomic Research Library? For example, a patient with a rare condition. How do we let the patient know about this?

**Natalie:** That's a really interesting and important part of our model here at Genomics England because we're not just a research library. A huge part of our mission is to enable researchers to discover something new that could potentially be a diagnosis for a participant, and for that actually to be fed back into the NHS and perhaps make a difference to their own care and treatment.

For a lot of our participants with rare conditions, you know, they don't have a diagnosis, they're spending years on this - we call it a diagnostic odyssey. You know, waiting and hoping for some insights and a diagnosis on their condition.

Actually, if researchers working within our research library find something that they think could be a new diagnosis and that that could apply to one of our participants, then we at Genomics England hold the key as it were to that person's identity. We can act as a facilitator. We can take that potential research finding, and if we think we're confident, it really could be a diagnosis for that individual, we can kind of facilitate to go back to the NHS, to that person's clinician and have that discussion with them to say, look, this is something research has found, they think it could maybe be an answer for this participant, what do you think? We have that discussion with the clinician and then they take responsibility then for going back to the participant if they think it really could be a new diagnosis for them.

**Naimah:** Okay, and finally, you just mentioned there about a diagnostic odyssey. Can you explain to me what that is?

**Natalie:** Sure. So, for a lot of people who have a rare condition, and there are lots of people with rare conditions, they are quite common. They may spend years going through tests, evaluations, you know, back and forth. Their symptoms might change and evolve over time, lots and lots of hospital appointments, and it's not always clear what these conditions are. And in genetics and genomics, you know, some of these conditions can take years to find out what they are. It can take a very long time to find a diagnosis and a lot of frustration, a real rollercoaster. And you know, if you just think about, for a moment, the anxiety and trauma that would entail, blood tests, appointments, back and forth. You're thinking about children who have to go through a really long, torturous process of finding a diagnosis. They're very hopeful and then they don't get an answer. Then they discover it's not the condition they thought it was. That can take a really long time and it be very emotionally challenging and draining.

Our hope is that by advocating for whole genome sequencing, we can shorten that odyssey, that period of time during which participants are looking for answers and just not finding them. Because with whole genome sequencing, we are hoping that we can really shorten the length of time it takes to find a diagnosis for those rare conditions.

That's really the hope and the promise of whole genome sequencing. Something that we hope that through the National Genomic Research Library, we really can make this difference to people's lives and give them the answers that they're desperately hoping for.

**Naimah:** Thanks, Natalie. That was Natalie Banner explaining what the National Genomic Research Library is and how it's used for health research.

If you've got any questions or any other concepts, you'd like us to explain in our Genomics 101 series, please contact us at info@genomicsengland.co.uk

Thank you for listening.