How can we unlock the potential of large health datasets?

Behind the Genes transcript

**Natalie: Welcome to Behind the Genes**

**Naomi:** So, we talked to each other quite regularly. We have tried to learn from each other about the efficiencies of what to do and what not to do in how to run these large-scale studies efficiently. When you are trying to recruit and engage hundreds of thousands of participants, you need to do things very cost effectively. How to send out web-based questionnaires to individuals, how to collect biological samples, how the make the data easily accessible to researchers so they know exactly what data they are using.

All of that we are learning from each other. You know, it is a work in progress all the time. In particular you know, how can we standardise our data so that researchers who are using all of us can then try and replicate their findings in a different population in the UK by using UK Biobank or Our Future Health.

**Natalie:** **My name is Natalie Banner, and I am Director of Ethics at Genomics England. On today’s episode we will be discussing how we can unlock the potential of large health datasets. By that I mean bringing together data on a massive scale, including for example genomic, clinical, biometric, imaging, and other health information from hundreds and thousands of participants, and making it available in a secure way for a wide range of research purposes over a long time period.**

**Through collaboration and industry partnerships, these programmes have the potential to transform research and deliver real world benefits for patients and health systems. But they also come with challenges ranging from issues in equity and ethics through to logistics, funding, and considerable technical complexities. If you enjoy today’s episode, we would love your support. Please like, share, and rate us on wherever you listen to your podcasts.**

**I’m delighted to be joined today by 3 fantastic experts to explore this topic. Dr Raghib Ali, Chief Medical Officer and Chief Investigator at Our Future Health. Professor Naomi Allen, Professor of Epidemiology at the Nuffield Department of Population Health, University of Oxford, and Chief Scientist for UK Biobank, and Dr Andrea Ramirez, Chief Data Officer at the All of Us Research Programme in the United States.**

**Andrea, if I could start with you. It would be really great to hear about All of Us, an incredibly ambitious programme in the US, and maybe some of the successes it has achieved so far.**

**Andrea:** Absolutely. Wonderful to be here with you and thank for you for the invitation. The All of Us Research Programme started in 2016 from the Precision Medicine Initiative and was funded with the goal of recruiting 1 million or more participants into a health database. That includes information not only from things like biospecimens including their whole genome sequence, but also surveys that participants provide, and importantly linking electronic health record information and other public data that is available, to create a large database that researchers that access and use to study precision health.

We have recruited over 830,000 participants to date and are currently sharing available data on over 600,000. So, we’re excited to be with your audience, and I hope we can learn more and contribute to educating people listening about precision medicine.

**Natalie:** **Thank you, Andrea. And not that this is competitive at all, but Raghib, as we are recording this, I understand the Our Future Health programme is marking quite a phenomenal milestone of 1 million participants. Would you mind telling us a little bit about the programme and something that you see as the benefits of working at scale for health research.**

**Raghib:** Thank you very much. So, Our Future Health is a relatively new project. It was launched in 2020 with the aim of understanding better ways to detect disease as early as possible, predict disease, and intervene early to prevent common chronic diseases. Similar to All of Us, we are creating a very large database of participants who contribute their questionnaire data, physical data, genetic data, and linkage to healthcare records, with the aim as I said, to really improve our understanding of how best to prevent common chronic diseases.

So, we launched recruitment in October 2022. Our aim is to recruit 5 million participants altogether, and in the last 2 years about 1.85 million people have now consented to join the project. But you are right, as of last week we have what we call 1 million full participants, so people that have donated a blood sample, completed the questionnaire, and consented to link to their healthcare records. In our trusted research environment, we now have data on over 1million people available for researchers to use.

Of course, we have learnt a lot from the approach of UK Biobank, which we are going to hear about shortly, but the resource is open to researchers across the world, from academia, from the NHS, from industry, so that will hopefully maximise the benefits of that data to researchers, but as I say with a particular focus on early detection, early intervention, and prevention research.

**Natalie: Thank you Raghib. Great to have you with us. Naomi, Raghib mentioned that UK Biobank has been running for a long time, since 2006. It is a real success story in terms of driving a huge range of valuable research efforts. Could you talk to us a little bit about the study and its history and what you have learned so far about the sort of benefits and some of the challenges of being able to bring lots of different datatypes together for research purposes?**

**Naomi:** Yeah, sure. So, UK Biobank started recruiting 0.5 million participants in 2006 to 2010 from all across the UK with a view to generating a very deep dataset. So, we have collected information on their lifestyle, a whole range of physical measures. We collected biological samples, so we have data on their genomics and other biomarkers. Crucially because they recruited 15+ years ago, we have been able to follow up their health over time to find out what happens to their health by linkage to electronic healthcare records. So, we already have 8,000 women with breast cancer in the resource, cardiovascular disease, diabetes, and so on.

But perhaps most importantly, not only does it have great data depth, and data breadth, and the longitudinal aspect, is the data is easily accessible to researchers both from academia and industry, and we already have 18,000 researchers actively using the data as we speak, and over 12,000 publications already generating scientific discoveries from the resource.

**Natalie: So, we have got 3 quite different approaches. Recruiting in different ways, different scale, different depth of data collection and analysis, but all very much around this ethos of bringing lots of different datatypes together for research purposes. I wonder if you could talk a little bit about how you might be sort of working together, even though you have got slightly different approaches. Are there things that you are learning from one another, from these different data infrastructures, or how might you be looking in the future to work together to address some of the challenges that might come up from working at scale?**

**Naomi:** So, we talk to each other quite regularly. We have tried to learn from each other about the efficiencies of what to do and what not to do in how to run these largescale studies efficiently. When you are trying to recruit and engage hundreds of thousands of participants, you need to do things very cost effectively. How to send out web-based questionnaires to individuals, how to collect biological samples, how to make the data easily accessible to researchers so they know exactly what data they are using.

All of that we are learning from each other, and you know it is a work in progress all the time. In particular, how can we standardise our data so that researchers who say are using All of Us can then try and replicate their findings in a different population in the UK by using UK Biobank or Our Future Health. So, can we come up with common standards so that researchers can better directly compare the data that they are using? So, we are in close contact with each other.

**Natalie: Fantastic, thank you. And Andrea, from your perspective obviously you are collecting data in the US. Are you finding ways of working internationally and with other infrastructures like Biobank and Our Future Health around things like data standards? It sounds like something simple, but I can imagine it is quite complex in practice.**

**Andrea:** Absolutely, and that dialogue and understanding and learning from each other both informally in meetings and talking as well through the published literature. So, all of these datasets are actively widely used, and seeing what is coming out in publications helps us know what researchers are doing with the data. And when you see different researchers either generating hypotheses from our datasets in a different way, or testing hypotheses differently, that helps us understand where some benefit might be added to our dataset or where we really may need to grow in a different direction to meet some other research needs.

I think that every study design always struggles with that balance between knowing exactly what we want to study and therefore building very specific questions and very specific protocols, but also allowing for the knowledge that we don’t really know all of the discovery we need to make and bringing in datapoints that will really generate those new hypotheses for the future.

I think for our study in particular, UK Biobank has been so remarkable in this way, helping structure All of Us to be able to contact our participants like UK Biobank and say, “Hey, we didn’t really know what we were going to get, but we have put all this wonderful data together and now we need to do a deeper dive.”

So, the engagement and long-term return of those UK Biobank participants has really enriched our data, and we have learnt from UK Biobank a lot there, and hope through growing our partnerships programme that we can continue to create partnered research opportunities to strengthen that data as well. That is a new thing coming out of our group. You may have heard of it previously as ancillary studies, but we recognise the partnership that is important for those research opportunities. So, we are reporting here that we are hoping to rebrand it to reach a larger audience, and that is led by Dr. Shelley, as partnered research opportunities that will allow us to re-contact, bring our participants back, and really deepen that dataset.

**Natalie: Thank you. And Raghib, I know that it is a really important part of the Our Future Health model about going back to participants, but you are in quite early stages of working out what those opportunities might look like.**

**Raghib:** Yes, very much early stages. Just to reiterate the point for me personally, having started my research in the UK about 20 years ago, I have certainly learnt a lot personally, but we have all learnt a lot from the model that UK Biobank established in terms of collecting data and providing it to researchers, and I see these 3 studies as very much complimentary.

All of Us again have done a lot more work in terms of providing feedback to participants about their risk of disease and genetic information, and as you say Our Future Health was set up deliberately to not just be a purely observational study, but to give participants feedback about their risk of different chronic diseases as well as the opportunity to take part in not just studies to collect data, but also interventional studies to see if we can change the natural history of disease and prevent diseases in our participants.

So, that has never really been done at scale before, and that is certainly a big challenge for us to do, not just in the UK, but anywhere, including the US and working with health systems as to how best to do that. So, you know we have spent the last 2 years really trying to understand how best to recruit participants and to provide data to researchers for the next couple of years, and long beyond that we will be looking really as to how we can maximise the benefits of providing feedback to participants and taking part in interventional studies.

**Naomi:** I think one way in which we can all learn from each other actually, is we know how to recruit hundreds of thousands of people, the general population, into research study, and the next challenge is how do you keep engaging them, telling them what you are doing. You can't collect everything when they first join the study, or they would be with you for days. So, what UK Biobank has been doing is sending out web-based questionnaires, a couple a year, to find out extra information about health outcomes, lifestyle factors. Inviting them back to specific assessment centres.

So, we are inviting 100,000 participants back for imaging, and then again over the next few years for a second scan. So, I think the real challenge here is once you have recruited them, how to find that right cadence of engaging those participants to keep contributing their data and their biological samples to really maximise the value of the dataset for research. That is an ongoing challenge for all of us. But I have to say, the UK Biobank participants, they are an amazing group of individuals, very altruistic.

Our Future Health and All of Us, we don’t give feedback, so there is nothing in it for our participants other than knowing that their data may help the future health of their children, and their grandchildren, and the rest of the world. So, that is very humbling, to know that the data that they have generated, and we have collected on them, is being used in that way.

**Natalie: That’s a really interesting point, Naomi, about the difference between a research study that is designed for answering a particular question. You gather specific data for a specific purpose, and when it comes to recruiting participants into that you can be very clear about what it is you are trying to do.**

**But of course, for all of these programmes, the whole nature of them is that you are collecting a lot of data over a long period of time, and it could be used for all sorts of different purposes. You can't say at the outset exactly what those purposes might be and what those outcomes might be. So, there is a really interesting question, and of course I would say this with my ethics hat on, a really interesting question around sort of participant trust and confidence in those programmes.**

**Naomi, you spoke just then about one way of retaining engagement and retaining people’s interest, but I wonder Raghib and Andrea, if you have got thoughts on those sort of questions of how you can create that environment where participants can trust what you are doing with data over a long period of time, when you can't at the point at which they consent, say exactly how that data might be used? You have got a sense of the kinds of purposes, but you can't be too specific**

**Andrea:** Sure. We know, and I have learnt from my own peers in this role, that enrolment in the study isn’t the end point of engagement. All of Us’s approach on engagement has been communicating with the entire community and really being there in the community, and that has been very powerful.

One effort over the last year we are proud of has been what we are dubbing participant driven enquiry, and that is where we say, “Thank you participants. We have gotten a ton of data out there for use, and funded researchers to use it all the time, but what do you, the participants, really want?” We were able to then take papers that researchers write and help tell participants and explain it in lay language, so the participants can say, “Hey, I have a question. Could you answer that for me?” Maybe we can, maybe we can't, but it has been very interesting to hear what participants want to know, and that participant driven enquiry project has turned out to be a big opportunity there.

The question they came to was not easy. Certainly, we didn’t expect an easy question, but they came to us asking, “Why is my diabetes worse than someone else’s? Is it the environment? Is it my genome? Is it my access to care? Why can't my diabetes be as well controlled as someone else’s?” So, that has been huge, to interact directly with our participants and help really close the loop by answering questions in the language of research and show them how their data is contributing back.

**Natalie: Thank you. And Raghib, how are you sort of grappling with these questions, particularly because you are recruiting so very heavily at the moment?**

**Raghib:** So, as you say it is a challenge, and people do join the programme primarily based on trust that we will use their data for public health benefit and for the benefit of the whole population, but they also join on the basis that they will get back information about their own health and their risk of disease. To do both of those is not straightforward. I mean, the first of those, it has been well established by UK Biobank, and about 80% of our participants also say they are doing it primarily for to altruistic reasons, which is great. But 80% also said they would like to receive feedback about their own health, which is also understandable, and so we need to find ways to provide that in a timely way, but also in a way that the health service can manage. That is going to be one of our key challenges going forward.

But to echo what Naomi and Andrea have said, I mean to maintain participant’s engagement with the programme is not easy. We need to make sure that they are receiving information regularly, are kept up to date with what we are doing with their data, with the work that we are doing with academia, with the NHS, with industry etc. It is easier now than it was before because Our Future Health has been set up as a digital cohort, so we have means of communicating much more easily with our participants. But yeah, as you say we are at early stages. Over time that does get harder, to maintain that engagement. So, we know in the next one to 2 years we need to step up our work on feedback and recontact.

**Natalie: Fantastic. I really love the idea of like the participant-led enquiry. That is something that I think our participant panel at Genomics England would really like to hear more about.**

So, speaking about sort of ongoing engagement with participants, one of the challenges we know around recruiting into largescale studies like this is that many research datasets don’t have equal representation from all communities. That might have an impact on the quality, the representativeness of the scientific outputs that you can generate, and potentially the benefits back to patients and participants.

How are you addressing this challenge in recruitment where you may have some communities that are not as engaged with scientific research. You may have elements of distrust or people being marginalized, having difficulty accessing research and these sorts of opportunities. Do you have any examples of what has worked really well? Raghib, if I could come to you first.

**Raghib:** Sure. So, I mentioned I worked on UK Biobank about 20 years ago. One of the things I was looking at then was how we could maximise participation, particularly of people from ethnic minorities into the project. Because of the age group that was chosen by UK Biobank for very good reasons, age 40 to 69, the proportion of people from ethnic minorities was relatively small. So, although it was representative for that age group, I think it was about 6%, or 34,000 out of the 500,000, that were from non-white ethnic minorities.

So, when Our Future Health was set up, we knew that the population has changed anyway. You know, the UK has become a much more ethnically diverse society. But also, because it is a cohort from 18+ and I think minorities tend to be younger on average than the white population, we knew we had an opportunity to really have a big step change in the number of people that could take part in a study like this. So, our aim is actually to get 10% of the whole cohort from ethnic minorities, so 500,000 out of the 5 million from ethnic minorities. Actually, so far we are pretty much on track. So, of the 1.8 million that have consented, about 180,000 are from non-white ethnic minorities.

That is extremely important, particularly for genetic research where non-European populations are very much underrepresented in nearly all genetic databases. Secondly, from a UK context, although it applies of course in all countries, is that people from more deprived backgrounds are also less likely to take part in this type of research. So again, we have made a very deliberate attempt to try and ensure we have adequate numbers from the most deprived quintile. Again, about 10% of the cohort so far, nearly 200,000 are from that most deprived quintile who both are underrepresented in research, but also have the worst outcomes. So, this is really our first study that has been big enough in the UK to look at that group properly and understand some of the factors at an individual level that we haven’t been able to in the past.

Finally, geographically, so the first time again because it is a digital cohort, we were able to recruit people from all over the UK. So, every single part of the UK is now represented in Our Future Health, particularly coastal communities and rural areas that haven’t been able to take part in this type of study before, as well as Northern Ireland. You know, for the first time we have got that full geographical coverage.

**Natalie: Fantastic. I suppose a lot of that recruitment approach has very much been about going to where people are, rather than expecting them to come to you. Is that right?**

**Raghib:** That is right and thank you for reminding me. So yeah, we have had a different approach. So, we have opened up many, many more clinics than previous studies through a combination of mobile units, shopping centres, community pharmacy. Community pharmacy in particular has been very important. So, to date we have had about 400 different venues that we have been able to recruit. That is over 1 million people that have given blood samples, and that has really enabled people from every part of the country to take part. Secondly, we have kept clinics open in areas of greater deprivation and ethnic diversity much longer than in other areas, to maximise the opportunity for them to join. Thirdly, we do provide reimbursement for people with expenses to ensure they aren’t excluded because of financial reasons, and again that has helped.

**Natalie: So, really making those efforts is evidently paying off. Andrea, have you had similar experiences as All of Us? What has your approach been to try and ensure that you are getting a wider representation from different communities?**

**Andrea:** It has really been a focus on the programme from the start to engage those who have not been included in research in the past and make sure the opportunity is there to participate. Our Engagement Division, led by Dr. Corrine Watson has really pioneered reaching those communities here in the US.

I think one other thing I will mention that we think about when we think about how to engage participants and reach people to return value back to those communities, is to make sure the people who are accessing the data also represent them, and we can build diversity within that researcher workforce. So, since our data was first released in 2020, we have recognised that the biomedical workforce also has a huge group of underrepresented individuals, and a lot of our researcher engagement and researcher outreach has focused on reaching those of diverse backgrounds and career paths.

To that end we have reached out and engaged historically black colleges as well as other minority serving institutions, really looking to make sure that their students and researchers can have the same access as more traditional research-based institutions in the US system.

That has been important because our system is built on cloud-based architecture and shared data that doesn’t require a huge cluster on campus, and that helps remove a barrier that some of those institutions and researchers may have had. We also know they haven’t been able to participate in the past, and we think that cloud architecture again can make the data much more feasible and be a huge support to diversifying the researcher workforce as we go forward. That circling back, helping them be the voices speaking to their community, helps build out that diverse participant community base as well.

**Natalie: That’s such an important point, because it is not just about the participants and the data you can collect, but also who is able to look at it? Who is actually able to undertake the research?**

**Naomi, can I bring you in here? I know that UK Biobank has been thinking a lot about researcher access to data and trying to ensure that the data that you hold, the really rich datasets you hold in UK Biobank, are more accessible to researchers from different backgrounds who may not have the same level of resources. Can you tell us a little bit about the work you have been doing on that?**

**Naomi:** Yeah. So, just following on from what Andrea said, it is really important to get as diverse ideas as possible from across the global research community to really move public health forward.

So, what UK Biobank has done is we are putting mechanisms in place so that early career students, and career researchers, and researchers at all levels of their career from lower income countries, can access the data at a much lower fee. So, currently for most researchers it costs about £9,000 to access all of the data. So, that is 40 petabytes of genomic data, biomarkers, clinical outcomes, lifestyle factors and so on. So, early career researchers and those in lower income countries, it is about £500.

On top of that a group of big pharmaceutical companies have got together to create a global researcher access fund, which essentially covers this reduced fee so that all researchers no matter where they are from have exactly the same opportunity to access the data to advanced scientific discoveries. So, on top of that all our researchers now use our online secure research analysis platform. While there is no charge to access the platform, there are costs associated with compute needed to analyse and store the results.

So, AWS have donated research credits for early career researchers and those from lower income countries up to a total of about $500,000 per year, to use the research platform. So, researchers can apply to use these research credits to offset the costs of compute and storage. So, that means that we are trying to democratize access to researchers from all around the world.

I think actually our biggest challenge is not so much … we have largely dealt with you know subsidising the cost. It is actually making researchers from lower income countries aware that these resources exist, and that are applicable to them.

So, sometimes we hear from say researchers in Africa or South America, “Well, there is no point accessing UK Biobank because it is not relevant to our population.” You know, a third of our researchers are from China. So, even if UK Biobank hasn’t got coverage of those racial ethnic populations, that doesn’t mean that the associations that you find between risk factors and disease risk are not applicable to other different populations. And that is also why having different resources like UK Biobank, like Our Future Health, like All of Us, in different populations around the world, is so important in order to replicate those findings.

**Natalie: Absolutely, and fantastic just to hear the attention that is being paid to trying to ensure that diversity of different types of researchers who will just bring different questions to the table, different perspectives on the data, different priorities, different types of questions.**

**So, speaking about that diversity of researchers, one really important part of his ecosystem that we haven’t really touched on so far is around the role of industry. There are a lot of really important research questions being addressed by industry. Some that can only really come from, maybe it is pharmaceuticals, maybe it is tech.**

**From your perspectives, what kind of role can and should industry and commercial partners play in supporting the kinds of long-term research studies that you have set up, and ultimately trying to get to that point of sort of generating benefits back to patients and health systems. Naomi, can I start with you, for that sort of longer-term perspective for Biobank?**

**Naomi:** So, industry are great partners for long-term studies like ours because they can bring additional funding, expertise, and technology. So, for UK Biobank, because it is so easily accessible to industry and academics alike on exactly the same terms, what it has meant is that industry, particularly big pharma and also now big tech, they can access the data, they see the value of the data for their own research purposes, and then they have invested into UK Biobank to do whole-exome sequencing, whole genome sequencing, proteomics at scale to increase the value of the dataset for their own drug discovery pipelines.

But of course, it means that the data that they have generated, which cost millions of dollars to generate, when you need deep pockets to do these kinds of study enhancements, then become available to all researchers. So, having access to these largescale resources that have deep data on genomics, physical measures, other biomarkers, and clinical outcomes enables pharma to rapidly increase their drug discovery pipelines in generating new drugs and treatments for patients, and also those data are then shared with the rest of the global research community.

So, we found it to be a really exciting win/win in which industry get what they need to help move forward new drug targets and discovery, but also other researchers get what they need in order to make other scientific discoveries in different fields of research.

**Natalie: Thank you. And Raghib, I know that for Our Future Health, that industry relationship is a really important part of the founding model. Will you tell us a little bit about how you are engaging and working with industry partners?**

**Raghib:** Sure. So, as you said Our Future Health was set up in a different way, as a very public private partnership. Although the largest funder is the UK Government, more than half of our funding has come from a combination of life science companies, so pharmaceutical, diagnostic companies, as well as the medical charities, so the larger medical charities in the UK. That partnership is deliberate for all the reasons that Naomi has outlined. There are areas where academia and the NHS are very strong, and areas where industry is very strong, and by working together as we saw very good examples during the pandemic with the vaccine and diagnostic tests etc, that collaboration between the NHS and academia industry leads to much more rapid and wider benefits for our patients and hopefully in the future for the population as a whole in terms of early detection and prevention of disease. So, we have 16 life sciences companies that have joined as founding partners with Our Future Health who have contributed financially to the programme.

Equally importantly they have also contributed scientifically, so there is a huge amount of scientific expertise in industry, and they work with us with our Scientific Advisory Board with our scientists internally to think about the best use of the resource for drug discovery, diagnostics, new medical technologies, and new targets etc.

So, that is the vision, and so far, it is working well. It is a relatively new model to have set up a project like this in this way, but it has been a very collaborative approach, and we all recognise, all have similar aims, so recognise what we are working towards. You know, we meet regularly. We have a Joint Founders Board where as I say academia, NHS, industry, and the charities come together to decide on the priorities for the coming years.

**Natalie: Fantastic. And Andrea, I suppose in the US it might be slightly different culturally from the UK, but the role of industry with All of Us, how are you engaging with those pharmaceutical, technology bodies, and partners as well?**

**Andrea:** Absolutely, and maybe this goes back a bit to your first question. We at All of Us love learning from UK Biobank and have really seen them forge a lot of wonderful partnerships that have enriched and developed their dataset. We at All of Us have started with academia and working through partnership opportunities really intramurally at intramural centres that make up parts of the National Institute of Health. We believe that building on those close friends and family relationships we have both in the government and academia get us through our first step to be able to interface with commercial organisations. That really started with taking the first step this year to ensure broad availability of data that can maximise both use of the data available, as well as look forward to our partnership opportunities in the future.

So, commercial organisations as of 2024 have also been able to access the All of Us dataset that is that first step in thinking about what a partnership would be, and we are glad to build on the access that international organisations and academic organisations already have.

**Natalie: A lot to look forward to here. We are going to have to wrap up in a moment, so I’d just like to leave you all with a final question before we have to end the podcast. There is huge ambition in all of the research programmes that you are leading and involved in, but what are you most excited about coming down the line in the next few years? What do you think is going to be feasible? What really gets you excited about the work that you are doing and where you see the potential benefits really landing in the next few years? Andrea, would you like to start?**

**Andrea:** Thanks. There is a lot we are really excited about. I haven’t had a chance yet to mention our paediatric cohort, and that in addition to expanding access for international research, in 2024 we were able to enrol our first paediatric participants. That really sets up the potential to observe participants across the lifespan. That is a huge advance for All of Us and we are excited about the paediatric work going forward.

**Natalie: I love that, how do you come into the future with us? That is fantastic. Naomi.**

**Naomi:** Yeah, if I had to choose one would be the possibility of being able to measure circulating proteins on all half a million participants. We have done this on about 55,000 participants, and just that subset alone is already generating fascinating insights for early biomarkers for disease through protein profiles and risk prediction of disease. I think having that on all half a million coupled with their genomics data and health outcomes, will bring a sea change in how we diagnose disease earlier. So, I think that is a really exciting avenue for us to go into over the next couple of years.

**Natalie: Really enriching. That data sounds like a very exciting set of possibilities. Raghib.**

**Raghib:** Thank you. There are so many opportunities here, but I will just maybe mention 3. So, the first, in terms of being able to combine the genetic data that we are collecting and all the other information about risk factors, and particularly the fact that we have this on a lot of young people, will enable us to identify people at high risk of diseases in the presymptomatic phase and then to be able to offer them both feedback about their risk of disease but also interventions that can change their natural incidences. That has never really been possible before. That is extremely important for all diseases for people, but also it is very important for our healthcare system.

So, those of you listening in the UK, I know the NHS is under a huge amount of pressure, and the current model of healthcare which has been in place really since the inception of the NHS, is to treat late-stage disease when people have already developed symptoms and signs. You know, it wasn’t really possible to identify people earlier, but it is now, and Our Future Health will provide the evidence base to show that prevention really is better than cure, and to show that these approaches both lead to better clinical outcomes, but also are cost effective and a good use of resources. Of course, the new government is very much committed to this as well, you know moving from acute care to prevention, from hospitals to community, and from analogue to digital.

Finally, because our cohort has now become so large and does cover every part of the UK, and this wasn’t something I necessarily thought about when we started Our Future Health, we are able to have unique insights into the health of the population across every age group, across every ethnic group, across every geographical area, and by deprivation, and to understand not just observationally in terms of risk factors, but also the impact of interventions on those different populations.

We can look at that, as I said at an individual level on millions of people to gain intelligence about what is going on in terms of public health, but also to see what will hopefully improve their health in the future. So, there are really, you know I have described transformational opportunities to improve health through both biomedical research and populational health insights now through the resource, and I look forward to working with colleagues across the UK and globally to deliver them.

**Natalie: We will wrap up there. Thank you so much to our guests, Dr Raghib Ali, Professor Naomi Allen, and Dr Andrea Ramirez for joining me today as we discussed how collaboration, scale, ongoing engagement, can really unlock the potential of largescale health datasets to drive brilliant new research and ultimately improve the lives of patients and the population.**

**If you would like to hear more like this, please subscribe to Behind the Genes on your favourite podcast app. Thank you for listening. I have been your host, Natalie Banner. This podcast was edited by Bill Griffin at Ventoux Digital and produced by Naimah Callachand.**