Aris Baras JULY 2018 CRI - PRESENTATION - 17 MINS
[MALE RESPONDENT]
[Other comments:]

Thanks to Sekar for taking us through a good example about having this population scale of genome typing data available allows us to get a better understanding of some of the drivers of disease. Moving swiftly on, I'm delighted to welcome to the stage Dr Aris Baras, vice president and head of the Regeneron Genetic Centre to outline the incredible work that the team are doing, and to share his perspectives on the next phase, and the introduction of exomes data to the Biobank resource, Aris.

That was amazing stuff, thank you Seg. It's always fun trying to follow Seg. I always feel like I'm going to follow Seg. All right, well, it's a real pleasure to be here. A few things I wanted to say that really echo a lot of what you heard today, and a huge thanks to Rory and Mark and Naomi, and everyone we've heard about at UK Biobank. I mean this is truly exceptional, right? We're all keenly aware how hard this is. What they've done over the years to put this resource together, we've had the great privilege to be involved with them on this one initiative in terms of exomes sequencing, and it is extraordinarily hard to pull these things off, and they do it so gracefully and they do it in a way that they make the resource open to everyone, so a great thanks to you all. We are really privileged to be working with you.

We were asked about really the value of exomes sequencing and what we can add as a group here to the resource, and to all of you, and what we can do in terms of new genomic discovery and impacting science and medicine. You really hopefully are convinced at this point and ought to believe that really the tipping point here, I kind of see it as equivalent to the semi-conductor industry taking off, and where we've had a few millions of people who have been sequenced around the world, I think for here in five years or ten years talking about this we're talking about hundreds of millions of people that have been sequenced, whether it's broader screening and looking at monogenic disorders, use of these polygenic risk scores where otherwise you wouldn't be able to detect these kind of silent, if you will, risk factors to all the types of new insights, maybe innovations that we can bring forward because of this data.

The things that we're so excited about exomes sequencing I think, really everyone in the field who is working on sequence data, interpreting outside of the genes of something that will come, but right now one of the easiest things to interpret is really the highest value part of the genome in terms of genes and their functions, and so we can really ascertain that completely, or near completely through sequencing, and in this case exomes sequencing. There have been - I will tell you a few stories just in our hands and with our collaborators - but there is just really a flood of information coming out in terms of some really impactful mutations, things like loss of function mutations. We can start to understand what genes and their functions are by building a code, a foundation of knowledge between genes, how they're disrupted and their effect on health and phenotypical consequences, and we can also past guideposts and start to do some fine mapping and look at maybe more severe effects or directionality in those regions by looking at coding variation. We can also start to look at things that now we have a challenge in the community where there is lots of information

about potentially pathogenic mutations, and to have a resource like this, and to be able to link those to health records and health information, to have a better assessment of pathogenicity and to allow the clinical community to use that is unbelievably important and valuable.

As Mark talked about, it's part of a really fantastic overall genomics initiative that UK Biobank has really championed in terms of getting a large amount of genome typing and computer data out there quickly, following now quickly with exomes and now you'll also hear following with genomes as well, so really, really tremendous.

One perspective I'd like to share, in contrast to all that optimism of how great genomics is and how many people will be sequenced and how actionable it will be, one of the challenges we have - and I'll bring you to a perspective from those trying to discover new medicines and bring them forward - it's unbelievably hard, one of the hardest things that we as a community, industry academics, everyone try to do and we fail most of the time and we spend an exorbitant amount of money doing that failure. I think I'm very glad that all of us have basically found that to be unacceptable in this day with science and technology, and we must do better, and so I'll tell you a few stories of how we really think genomics is the key to that. It's kind of the pinnacle of our strategy, and we think it's the single greatest hope in terms of tomorrow's medicines and hopeful cures.

A few stories that I'll tell you about from our work, and really hundreds of scientists and collaboratives we've been working on with these, and some kind from hypothesis driven approaches. We have some opportunities where we have drugs in development, and we can ask questions in terms of mutations that might mimic their effects, and can we predict the effects of clinical trials, or use them to guide where we might start hoping to run trials or develop these therapeutics, as well as unlocking and unravelling new biology and new fields in places like chronic liver diseases where really there are no effective treatments for some of these conditions?

The first is talking about a programme that Regeneron is working on now, the biology of [?angeoponant 0:05:26.6] and angeoponant like genes for a few decades now, and then some of the seminal work done and understanding the role of angeoponant like genes and their products, and lipid metabolism was done by folks like Sake who you just heard from and Jonathon Cohen and Helen Hobbs who really established that from family and population studies. There is an absolute huge role in this gene and this pathway and lipid metabolism, and you see, when you disrupt this gene, and also in functional modelling, a reduction of all lipids. You see a reduction of LDL, which is good. You see a reduction of HDL, which was unclear, and you see a reduction in trichlorides, which is exciting and might be a new area beyond LDL lowering, but it was completely unclear whether this would have an effect, a beneficial effect in terms of events and outcomes.

We and other companies are involved in developing therapeutics and inhibitors of this pathway, and of this target in specific, and this took two or three years to pull together all of this data, but to look at loss of function mutations in this gene and to ask the simple question; will there be protection from cardiovascular disease risk? Will it be neutral? Might it actually go the other way? Ultimately, the story was that loss of

function is protective, and Sake and colleagues were also involved in this and have shown some more data, and this is really exciting and unbelievable capability as we're able to think about our targets and use information like this to guide effective development of these programmes. As you fast-forward to very near term here in UK Biobank, what took us two to three years to pull together all the data could be done in minutes when all the sequence data is available in this database.

Another story that I told you about was switching gears and looking at diseases where we are really struggling, and don't have many experimental therapeutics, let alone effective therapeutics that are there in practice. We looked at fatty liver disease and alcoholic liver diseases, non-alcoholic liver disease and these are leading causes of liver failure, sclerosis, liver cancers and we had an opportunity to look at a very large data set in our guys in our collaboration where we've sequenced about 100,000 people and to look for new genes involved in chronic liver disease. We were very fortunate to find a mutation that is a protein altering mutation, and ends up in a truncation of this enzyme, and leads to a pretty dramatic protection. You can see the heterozygotes and homozygotes have pretty profound protection from developing these forms of disease, including advanced forms of sclerosis and liver cancer. This is something that we're very excited to be actively working with Alnylam on developing an RNA inhibitor of this, a knockdown, and we have, just from this exciting information, probably about 50 people in our company and with collaborators now working out the biology over the last year-and-a-half and trying to bring therapeutics to patients as quick as possible.

I can't stress just how transformative this type of information can be when we feel like we're in the dark failing all the time, using suboptimal information and models to base programmes, and now you can have information like this to really guide drug development opportunities.

One other example and then I'll give you an update on the exomes sequencing project and UK Biobank. Another programme that we have much earlier, that we moved into early clinical development, and a lot of known biology about this interlocking this immune molecule, and a topic, allergic conditions, also respiratory diseases where A2P is a big component of them, and years ago we had some insights about loss of function mutations having potentially a protective effect on some conditions of great interest, and what I'm also showing here is not a genome mind polygenic score, but we had a genetic pathway score, and we were looking at scores of risk mutations in this target in its receptor, so in its pathway. You can nicely see that as you increase the number of burden of risk mutations, we saw an increase in risk and then we get a hold of the UK Biobank data, which fortunately had this loss of function variant, or a main loss of function variant was in that data set, and you can see how clear the picture is with ten times the sample size where you see this score gives you a dose dependent effect. Those in the highest risk score have 60 per cent higher odds of having disease. That is quite significant. Also, those with the loss of function mutation in the other direction have a pretty significant risk reduction.

We, and quite a few other companies are in the clinic developing an inhibitor against this, but we have also lots of emerging data around other conditions related respiratory and allergic conditions where we can now tailor and guide the development of this programme entirely based on genetics, and that is exactly what

we're doing.

This whole field of looking at loss of function mutations and human knockouts is incredibly important, and really helps us understand gene function. This is a little bit older data, but we did a pretty comprehensive analysis of just counts of loss of function mutations, all loss of function mutations or rare loss of function mutations, which most are rare. We looked at 215,000 exomes and we are very pleased to see that at this sample size we see most genes have at least one observation of a loss of function mutation, and we're starting to see a lot of genes have interesting numbers, at least 100 individuals with a loss of function mutation in that gene, and we can really start to test for changes in laboratory abnormalities or in risk of disease, phenolamide essentially. These curves, they are not plateauing. They will continue to shift up. That's what we are seeing and his would be an unbelievable resource to really connect and understand gene function at genetic level.

Talking about our exomes sequencing project, this started last year and really committed with a large consortium to the whole 500k at the end of 2017, so as much of a thank you to the UK Biobank team, but also especially to a lot of our colleagues who are making substantial investments to generate this resource and make it part of the resource for all researchers. We just completed, as Mark had mentioned, the first 50,000, so that was completed just about a couple of weeks ago, so our team in the last week or so tried to do some preliminary analysis, not around the phenotypes, but just about describing the data set and the genomic variation, and I'll show you just a snippet of that. We went from kind of whiteboard last Friday to some slides that Colin would have passed on and I pass them on to you all. The timelines are the following, so we're aggressively moving forward. Again, this is hard, but this is an amazing and highly competent team that has done this over and over again, and UK Biobank has really got a tremendous schedule to get us these samples and we should be able to get about 150,000 plus exomes done by the end this year, so we have started on that phase. Next year it's going to be an even bigger year, and we're really hoping we can get, and should be able to the whole 500,000 done by the end of next year.

Just a little bit of information on how the first 50k were selected, and we did this GlaxoSmithKline, and really we're thinking about the highest value density of data, a lot of the quantitative traits, all of the individuals who had been imaged at the time and the data was available, and so on the bottom right I have provided a cohort profile, if you will, but from the standpoint of what phenotype data is available in this data set. Otherwise it doesn't have any ascertainment bias or selection other than these three criteria of trying to get the highest density of information available on this first 50k.

The few slides that I'll show here from this data, it's as one would predict, no surprises, but we looked at basically a generic variance survey in the 50,000 exomes, and also compared this to a matched 50,000 individuals with the imputed data set, so you can make some fair comparisons. We looked at something, which we just talked about, and this pattern holds obviously in larger or in other categories of genomic variation. Obviously, exomes sequencing can't get too much beyond the coding regions and the buffer regions, but looking at loss of function variation, obviously you see a large amount, near complete ascertainment of the exome, and so you can see really kind of all loss of function events that are there. From the imputed data it

does a good job of picking up a substantial number of loss of functional variance, but since most of these loss of function variants are very rare, an imputation can really tap out when you get to the rare low frequencies. The point is that sequencing data, coming from the exomes sequencing data and also following with the genome sequencing data, is going to fill in a lot of important information, a huge amount of important information to this resource.

We also tried to quantify this a little bit. When you think about something like loss of function variants and how much power do we have to detect some changes, or the consequences of those? Based on some of that data I just showed you, now looking at how many genes have higher case carrier counts in the exome data, all as you would expect, or how much are ten times a higher count, so really significantly more powered in this initial data set that will all be available to folks for analysis for loss of function, for instance, versus what the current data set has. You can see a comparison on the left of exomes versus a subset of matched imputed data, and you can see how many genes have higher count in the exome data for loss of function counts, real counts or counter counts, and even when you go to something like ten times the enrichment. There are some genes certainly in imputed data that have higher counts, and this could be because the target or performance of the exome sequencing isn't perfect, or you might have some imperfections in terms of imputation or in low frequency in both platforms.

Then you'll also see that by sheer size of the 500k you will have many genes, maybe some lower frequency variants that are kind of a dominant one and can really drive counts of that loss of function variant. Al 33 is a perfect example, and you will have great power, so there are lots of genes present in the data set that can benefit from this, and so we provide some of that information. As we get into the 500k fully sequenced, it's really going to be fantastic in terms of this class of variation, but really all exonic and splice region variation.

A huge thank you to all of our team back at Regeneron, as I've already said. A million thanks to the UK Biobank folks and all of our collaborators, and from Regeneron's perspective really, we've got a network of 60 or 70 collaborators around the world, and growing, and this genomics community is amazing. We're just so thrilled to be a part of it, and have benefited tremendously, so thank you to many, many people.

[END OF TRANSCRIPT]