So can I ask all the speakers to please come up and take a chair. Just trying to keep you active, and could those people who have questions to ask, please make yourself visible to those with microphones.

Audience Male: Yes, a question about the polygenic risk factor, it's a scientific question, there is an argument whether there are, I don't know, let's say handle it or something like that, genes [unclear word 0:01:03.6] there is a mechanistic connection to the phenotype or is it really relevant to have 10,000 variants set out statistically appeal. In the way that you present it, is there a big difference in the performance and really the ability to separate the high risk from the low risk if you use thousands of loci or only hundreds?

M: That's a good question, the short answer is yes, using, going from the top hundred or few hundred loci to tens or hundreds of thousands of variants, increases predictive performance. I want to say significantly, but nontrivially. So I can't remember the exact numbers but with our data, if you just used a set of a hundred or two hundred loci for coronary disease maybe relative risks are three or three-and-a-half, and then if you use many, many more variants across the gene, that relevant risks go up to, in an extreme tail, go up to four-and-a-half or five. So it does make a big difference to include lots of the genome for prediction, of course, there's a separate question about using these sorts of resources to understand biology, where I think the loci where you're clear about what's happening, are much more helpful.

Audience female: Peter, can I ask a question, to add on to that, in our hands, when examining polygenic risk scores in non-European populations, we've not necessarily found the same to be true, and that using just the [?GMY 0:02:36.9] significant or a smaller set of variants might suffer less from, in the translation to other populations. Have you seen the same thing?

M: Absolutely, yes.

Audience female: Hi, this is a question mainly for Jonathan. So to what extent, your numbers when you're looking at loss of function mutations and how many genes have loss of function mutations, to what extent did you filter those variants for things like [?end truncations 0:03:07.3] that don't undergo non [unclear word 0:03:09.4] decay, or variants that are on only on exons that aren't actually expressed, because when we look in Nomad, those numbers are actually a lot lower when we do any kind of filtering on those variants.

M2: Yes, it totally agree, it really does matter. I think the numbers in Nomad are lower, I think one of the reasons we looked into this, several reasons. The precise definition of what counts as a loss and function variant matters, the filtering in terms of their, how deleterious they are, and also indels, so one thing that we include are indels, I think, Nomad less so, and that does change the numbers a lot. So I think this is a conversation that the community has to have. Is there really one canonical definition of a loss of function? Probably not, it's a sliding scale, whether you include indels or not makes a big difference.
Audience female: I've got another question for Peter about the polygenic risk scores and the low correlation with family history, and there are several possible contributing factors to that. I was wondering if you could just talk briefly about those and what you think's going on and also to what extent this differs between different diseases.

M: Good question, so we don't know the answer, but national hypothesis would be, and probably, many of them, family history is probably capturing to some extent, shared environment as well as shared genetics. Family history, so close relatives affected by disease will upweight the effect of rare variants in a way that polygenic risk scores won't, and polygenic risk scores certainly for some traits, maybe, we would hope less for disease traits, but could be capturing things other than direct genetic causes. They could be surrogates for social, socioeconomic markers, or nurturing variants and so on, where family history would be less relevant. There's also the obvious thing that polygenic risk scores are sort of less heritable than rare variants. So if I'm in the top one per cent, polygenic risk for a critical disease, the chance that one of my first-degree relatives is, it's ten per cent, it's not zero, but it's not massive.

Okay, we'll go to the back.

Audience male: Thank you, Nick Harvey from the MRC Life Course, epidemiology unit in Southampton. First of all, thank you very much to all for a fantastic talk. My question was really around the polygenic risk scores again, so I guess directed at Peter, we've been doing this sort of thing for bone outcomes with others and I was intrigued by the almost complete independence of the PRS from the QRisk and I just wondered what considerations there were in terms of the mechanisms identified by the PRS, relative to the treatment you might give. So how it might be used in clinical practice and I wondered whether the statin pathway or the cholesterol side of things was captured by the PRS, and whether people identified in that way, might be the best people to treat the statins and whether they are best identified through the queue risk and how those things come together in that distinction between maybe a relative risk reduction but at the lower end of absolute risk versus a reduction at the higher end of absolute risk.

M: Yes, it's a great question, so to put it another, how do we know that if you have a high polygenic risk score, but not necessarily and high traditional risk scores, statins would be a good thing. There's pretty encouraging data on that, so say [unclear name 0:06:55.0] has a really nice paper where he went back into the large, many of the large statin trials and where there was generic data available he broke the individuals in the trials up into five groups, so the bottom 20 per cent of PRS to the next and so on. That study showed that if you, amongst those groups, it was the top 20 per cent where the effective statins was largest in both relative and absolute risk reduction. So that's at a least encouraging evidence that those groups would benefit from statins.
Audience female: Thank you very much for the great talks, that's I will follow up on the polygenic risk score. I take the point that it's independent from Q risk, but at the same time the data you showed, showed clearly that the absolute risk difference in the groups is not higher. In other ways, the people that have Q risk below the threshold, despite the fact that their polygenic risk score was high, they were still below the threshold for treatment. So in a way that they're showing that the polygenic risk score doesn't add to the predictably of the Q risk.

M: Sorry, I may misunderstand the question, but I think it's clear from everyone else's and others have done similar things, that if you combine polygenic risk scores with traditional risk factors, you get improved risk prediction. So you get more people beyond thresholds in some sense you're spreading out your estimates of risk, because you can do that more precisely, and you get different people beyond the threshold. So there are individuals, were you to think that the threshold of ten per cent risk over ten years is in some sense an important threshold and certainly the UK guidelines for statin recommendations, if you think that's an important threshold, then there are, in our projections from UK Biobank to the UK populations as I said, there are half a million people who actually have that level of risk, that you're not aware of if you just use the traditional risk factors.

Audience female: Maybe it was not on the slide.

M: We have actually done the same thing in Iceland, looked at the polygenic score, combined with the traditional risk factor and we get out of it exactly the same results as Peter.

F: Yes, thank you.

Okay, we're going to number one.

Audience female: Hi, Jackie McArthur from. I've got a question for Kari when it comes to whole genome sequencing of UK Biobank, I was wondering what strategy you would recommend for sequencing this resource, based, short re-sequencing versus long re-sequencing based on the number of individuals and the cost involved and obviously the advantage when it comes to long re-sequencing and the interpretation of structural variation.

M: We have sequenced, whole GM sequenced about 80, a little bit of 80,000 people in Iceland, with short resequencing illumina technology. We have sequenced about 2000 people with technology, and we are
basically forced to use the short re-technology of sort of your basic sequencing because they're, at a rate they're still too high, the data management coming, although on technology it's too difficult still to use it in, as the short technology but there is no question about it. That reason we are sequencing with ONT technology is that it gives us much, much more sensitivity when it comes to picking up structural variants and we pick up probably about three times more structural variants using the ONT than a short re-technology. So if there was no resource limitation, my recommendation would be to use the short re-sequencing on all of the samples and then sequence the subset of with it with ONT technology. Something like 10 or 20,000, but as I said, that is all, that is our, that is the desire of an old man who sees no resource limitations anywhere where he goes.

**Thanks, okay, I think there's another one down here.**

Audience male: A question for David have you done the GWAS for penetrants of C282Y?

M: We're currently doing that in collaboration with the US study, the Hurst study and we've approached a French group. We've got some encouraging data, but we need to replicate it first before talking about it.

**I think there's yes, over there, number one.**

Audience male: I was interested in the announcement from John Bell on the early diagnosis initiative and just wanted further details on whether that was based around a research opportunity or principally around translation or both.

M: Yes, so it's going to be a research study, but it's going to be, has to by definition, be quite closely integrated with the NHS. So we'll be migrating data back to patients and into the health care system. So it sort of sits between the two. Obviously, we can't recruit those numbers of patients without using NHS systems and we will ultimately, I think, have our duty of care to return data to patients. So by definition there will have to be return of data in a variety of settings, but obviously, there's a whole set of issues, because it's a new type of cohort, where they'll be a dialogue between the study and the patients, but that's broadly what we're going try and do.

Okay, I think we'll stop there, because we have a fairly tight timescale for lunch. Lunch will be served upstairs where coffee was on arrival, you have exactly one hour, and I'm being told exactly one to eat. So on your marks, get set, and let's just thank the speakers from all of this morning's session again.

[END OF TRANSCRIPT]