You see I was born in the first half of last century, and I was brought up by parents who were socialists. One of the books that they used to read to me when I was a little boy, had to do with Robin Hood. The man who robbed from the rich and gave to the poor, and therefore, the Brits became my favourite nation. I admired them an awful lot, but that came to an end very swiftly in the early ’60s when the Brits invaded our fishing image, and I don’t think that the Brits realised how serious this was. I am now going to present to you some genetic data that support my contention that it could hardly have violated our rights more than invading a fishing image. It comes from a study of the genetics of perceptual [sic] smell. We actually, we found a variant in the gene that makes the trace-amine associated receptor type five. Which actually has a very significant impact on your ability to smell fish. So if you have this variant, the probability that you will be anosmic when it comes to the smell particularly of rotten fish. It's about with a ratio about three and it increases the probability that you will report the smell of fish as being pleasant. Actually it's interesting because in Iceland we have lived mostly on fish in all stages of fermentation for thousand years.

Actually, you remember this old saying, 'You are what you eat,' and actually frequently in the morning when I look at myself in the mirror I ask the question, 'Carry, are you getting old or are you turning into a fish?' Actually, when you look at a frequency of this allele it is 2.2 per cent in Iceland, 1.7 per cent in Sweden, 1.2 per cent in the UK Biobank and 0.2 per cent in Africans. So obviously we have been turning into a fish slowly and we are insensitive to our own smell. My love for the English was restored recently with the arrival of the UK Biobank which is the biggest gift ever given to science by any nation and eventually to healthcare and therefore helped by, I know you have been waiting for my forgiveness for a long time, but helped by, you are forgiven for invading our fishing limits. So this should be a good day for you. Then comes the question, why should the UK biobank be whole genome sequenced. So I am asked to express an opinion, I have been working on sort of Biobank related research for quarter of a century, I'm not asked to tell you about the results of that, but an opinion on whether we should or whole genome sequence the UK Biobank.

The reason it wasn't done in the beginning, the reason that they started with the whole exome sequencing, which has been done brilliantly by Regeneron and I am, as usual, very proud to the way in which they do things, but it is much cheaper, the sequencing throughput is tenfold great with the whole exome and whole genome. The instrument cost is less, but they have some time for some preparation is somewhat greater. A non-quotting part of the genome is about 98 per cent of the genome and actually we shouldn’t quote what is lost by whole exome sequencing, we should rather have it, what is not gained by whole exome sequencing, because what is gained by it is an awful lot and just working with the 50,000 that are already available has been very generous to us. There is also, there is a fraction of actions that are not captured by the whole exome sequencing kit and its regions are not amplified efficiently with PCR and that applies particularly to first exomes and the promoters and there is a capture of allele bias, and the more they advertise is similar to the reference, either more easily or efficiently is it captured and structural variations are less efficiently in the whole exome sequencing and this is just giving couple of examples of how their coverage is different and in the gene this is definitely because of richness in a part of the gene.
Actually, there is a recent examination of the loss or absence or loss of exomes to the whole exomes sequencing in about 7000 genes, there is a covering every exome, but in 12,000 to 13,000 there is a loss on average about 2.5 exomes. This is not a large number of exomes but this is still something that needs to be captured and we have at deCODE several discoveries that we have recently made, where the exome of importance was not captured by whole exome sequencing and we have some papers out for review describing that. In whole exome sequencing data, about half of the overlapping a variant have a wild type allele and half of them have the mutated allele, and the capture whole exome sequencing leads to systematic bias in this allelic balance and actually the allelic balance is one of the most important filters in whole genome sequencing for private mutations but it is much less so in whole exome sequencing because of the bias and the more different the haplotype is from the reference in the region of the capture probe, the more is their allelic balance bias.

Most of the common disease associations are with non-quoting sequence variations. We have an awful lot of things left to capture there, and the non-quoting variants will [unclear words 0:06:41.0] found but they are more challenging because of their numbers and difficulties for annotation, but if you only have good data on the exome sequence it will mean that you're going to lose a lot of positive variants and actually associations may be falsely ascribed to quoting variants because of disequilibrium with non-quoting variants. There are several papers recently out demonstrating the importance or having the non-quoting sequence, and I am not going to go into the details of those. Even in diseases like Autism, which we are most often looking at as rare variant diseases. The rare variant quoting mutation diseases, there is an enormous importance of their non-quoting sequence. The reason, one of the fundamental reasons that we need to sequence the UK Biobank is that it will become the largest reference in the world. Particularly important for clinical sequencing and the rich phenotype data will allow us to do all kinds of things and keep in mind that although the quoting variants are most likely to be associated with the disease, there are strong sequence variant annotations quoting of sequences that will also be enriched for associations and the enrichment of association gives direct assessment of biological relevance other than annotations. Having said that, it is important to realise that a few years from now, we will see that the reason to sequence their UK Biobank is totally different from the one that we are predicting today, or at least in my experience having worked on population resource like this for a quarter of a century, is that at least my guess, my prior assumptions, my hypothesis, turn out to be almost always wrong and the data humiliates me.

Anyway, we need to whole genome sequence the UK Biobank because they cover better the quoting regions, they allow us to understand associations better, they allow us to use understanding of their develop and understanding of non-quoting sequences much better, and actually the enormous amount of phenotypic data already in existence in the UK Biobank and the expectation of gathering even more, means that it is our duty to make sure that we have gathered as much data on diversity in the sequence as possible. I want to end by emphasising again the generosity that is reflected in the way in which the UK Biobank is run, providing everyone access. You guys have managed to do something that Americans have never been able
to do. There is not a single data resource in America that is as available in as generous of a manner as the UK Biobank, so congratulations guys, it was well done.

[END OF TRANSCRIPT]