Again I'd like to express my thanks to everybody who's taking part in UK Biobank, and all the researchers who have come here who are using it. UK Biobank was funded by the Medical Research Council and the Wellcome Trust. It was an enormous act of faith by them. The first ten years, which is what we are just coming to the end of, is really a time when UK Biobank is all work and no play, a lot of generation of data. Now the opportunities over the next five or ten years come to really understand the determents of a very wide range of disease. The whole idea of UK Biobank is to create a resource that many different researchers from all around the world can use to understand the determents of many different diseases. This idea of an open access resource for approved research from anywhere in the world, from academic and commercial researchers, in order that we understand better how to treat and prevent disease.

We actually started here in Manchester the first pilot study, to work out how to do things and how not to do things was in 2006. Then between 2006, 2010, with the support of the National Health Service we wrote to millions of people around the UK who live near to the assessment centres that we set up, and invited them to take part. Half a million did between 2006 and 2010 and many of you are here. Lots of questions, lots of measurements, lots of samples collected. As Mike Rawlins just said, with no return to individual participants taking part in terms of finding out about themselves, no feedback of any of the individual results. Chiefly because we don't know what they mean. We'll find lots of things, but the whole idea of UK Biobank is to find out what those things mean, how important are differences in the blood or, as we're now doing, differences in imaging. Consent for all kinds of different research, we're democratising research, we're putting data, huge amounts of data, really detailed data in the hands of lots of different researchers, coming from many different fields. Not just people who work in medicine, but people who work in informatics and statistics and mathematics, in finding patterns.

As I said, no preferential or exclusive access, it's a democratic process and one that perhaps works better than some of our more recent examples. We opened the resource up to researchers in the first quarter of 2012. As Naomi Allen, who will speak after me, will tell you, it's now being used by more and more researchers in more and more parts of the world, in more and more different kinds of very imaginative ways. The beauty of this is it's not constrained by the imagination of one or two researchers or research groups. It has the imagination of the world being applied to this huge database. After having provided £16,000,000 to set up UK Biobank for the Medical Research Council and the Wellcome Trust, the funders along with the Department of Health found some spare money in their back pocket. Suddenly we were given money to genotype all half a million participants, to measure 800,000 markets across the geno. Again, later in the morning you'll hear about how those data can be used in increasingly imaginative ways.

Just to talk about the genotyping, this was a three and a half year exercise that we're coming to the end of. We suddenly, as I say, got this offer of the funding to genotype UK Biobank participants [30 second silence in audio] where a company, Affymetrix, turned them into data. The data then came back to the

Wellcome Trust Centre of Human Genetics in Oxford. Peter Donnelly, Jonathan Marchini, Colin Freeman and their teams checked all those data, made sure that they were high quality and then used the data, the measured data to get even more information about the genotype. We released the data on the first 150,000 participants in May of 2015 and lots of different researchers have been using it. In fact, some 300 research groups around the world have got approval to use those data. So valuable are these data in terms of research, that no one wants anybody else to get it first.

We had this enormous problem at the beginning of this year of how do we get the updated data on all half a million people to everybody simultaneously, when it's the equivalent of millions and millions of filing drawers of data? Two petabytes of data, whatever that means, of data to all these 300 research groups have to be transferred simultaneously. They came up with the extremely interesting idea of sending the data, but it was encrypted, so you couldn't actually look at it. Then once everybody had downloaded it, which was going to take about a month, which we're at the end of now, we will send the code so that everybody can unencrypt it at the same time. Then it's a race to publish. Again, another lump of money just suddenly appeared. This is the thing, once you've got this resource, then they will come. The funders will come, many different funders will come and help to enrich it further and make it even more valuable for even more types of research. We've got funding, including from the British Heart Foundation, Diabetes UK, along with the main funders, to measure some 40 or so biomarkers in the urine, in the serum and in the red cells in the blood.

These are the different kinds of markets that we're measuring, some of them relate to cardiac disease of metabolic disease like diabetes, or bone and joint, or cancer, or renal, so kidney or liver disease. Those data are being generated at the moment. We released the urine measurements in November of 2016. Near the end of this year we're going to release what's called the hbA1c, which is like a time weighted average of your glucose levels. It gives you an idea about diabetes or diabetes control. Then about 30 or so measures in the serum, the lipids and things like that, around the beginning of next year. No one's done assays on this scale. The labs in Manchester, in Stockport have been working away, making sure they get very, very high quality data from all of these samples. One of the things we've realised is that if we want this resource to be available for the next 30 years, and people to be able to assay samples over the next 30 years, we have to use those samples very, very carefully. We have to ensure that they're used for assays that matter. Typically though when you have this kind of cohort what people say is, 'I'd really like to measure serum rhubarb in people who have got this disease and in people who haven't, and compare them and just do it in small subsets.'

The problem with that is that the assay methods can change over time, and so when you do an assay one time and then do the same assay a few years later, the results can be quite different. The assay method has changed, not because the people or their samples have changed. You get a lot of bitty data. If we want this resource to be useful for many different kinds of research, with those different studies being interested in the same thing that they're measuring, but to ensure that they're measured in exactly the same way with very high quality. It's much better if those assays can all be done at the same time on all the cohort. As is the case with

the genotyping, as is the type with the Bio markets that we're doing currently. Rather than taking bits of the sample and allowing researchers to take a bit, use it up over time, it is better if we just wait until we can afford to do assays on the whole cohort of particular analyse. This means that we can be much more efficient, we can ensure that when we do a pull a sample out of the freezer we measure lots and lots of things at the same time, so we don't deplete it.

We can improve the quality control. One of the things that we do is we pull the samples out in a random order, and then we measure the half a million samples in the random order, so we know that there must be no difference in whatever we're measuring. We can use the cohort as its own control of quality. We can make sure that there are no systematic differences in the assays. Of course, if we go to an assay supplier and say, 'Look, we'd like to measure half a million people rather than 50 people,' we get a very good deal. The genotyping of UK Biobank was probably about half the cost of what it would have been normally, because we were coming in with a bulk order, if you like. To do half a million assays, even on relatively inexpensive alights is expensive, the genotyping was over £20,000,000, the bio marker's £10,000,000. The question is, well, how do we do that with increasingly sophisticated assays? We've talked about genotyping, but you'll hear later in the morning about measuring the XM sequence, so measuring the sequence of the gene or whole geno sequencing, where you measure every single base pair. These are costs that are much greater.

Again, this is an incredibly valuable resource and one of the things that we're creating is a huge amount of data that's of interest, not only to academic but to commercial researchers. Again you'll hear about how a GSK/ Regeneron are funding the Exome sequencing, probably costs worth about $150,000,000. Those data will be in the resource for all researchers to use. Finally one of the big changes over the last few years is funding from the Medical Research Council, the Wellcome Trust, British Heart Foundation, Department of Health, to do imagining on 100,000 of the participants. We've imaged about 15,000 participants and in the afternoon you'll hear more about how those data are emerging, how they can be used. We've just opened the second imaging centre, the first one's in Manchester, the second is in New Castle, and that's now imaging. The third will be in Reading. Finally, everybody is linked in to their health outcome data, so death, cancer, hospitalisation data. For the first time ever in the UK, a big cohort linking into primary care data.

I'm not going to talk about the primary care data, Cathy Sudlow will talk about that later in the day. We think of ourselves, most of the time, as one nation. In fact, in terms of health records and the medical system, it's a devolved system. We have to go to England, to Wales, to Scotland to get these data, we have to go to different sources to get data on cause of death, on cancers, on hospitalisation. We have that all in place and every six months to a year we get updates of these data. They go into the database and then they're made available to researchers. We've just had our funding renewed by the Medical Research Council, the Wellcome Trust, the British Heart Foundation, and now Cancer Research UK have become core funders of UK Biobank. It's an expensive thing to keep on the road, but the more it's used, the cheaper it becomes, if you like. We're now looking at how we can get more information about health outcomes through linkage, then turning that

health record data into really reliable information about the health outcomes, [unclear word 0:14:04.1] typing, characterising the health outcomes in greater detail, working out how to manage this huge database.

Mark Effingham, our chief information officer, has this difficult problem of how do you make data on this scale easily available? Can we keep on shipping it out to researchers, or do we have to electronically ship the researchers into the data? Finishing off, these funded enhancements around genotyping, biochemical assays and imaging over the next five years. Then we're always thinking about additional things. Can we monitor the heart to look at cardiac arrhythmias? Is there something in people's stool that would tell us beyond blood and urine about disease? What are the next set of markers that we should be measuring in the blood? Biomarkers for infectious diseases, it's been said for many years that maybe infectious markers are related not just to infectious diseases, but to chronic diseases like cardiovascular disease, cancers and many others. Would measuring those markers be helpful in studying a range of different conditions? I'll hand over now to Naomi Allen, who is the senior Epidemiologist on UK Biobank, who's going to talk about the ways in which the resource is being used. I hope we'll have some time for some questions. Naomi.

[END OF TRANSCRIPT]