The UK Biobank
John Newton, Ph.D.

DR. TUCKSON: Now let me invite John Newton from the U.K. Biobank to share his perspectives. You've come a long way, so thank you.

DR. NEWTON: Thank you very much, Mr. Chairman, for inviting me. It has been very interesting listening to the previous speakers, and it is a pleasure to be here to tell you more about the U.K. Biobank.

The first thing to say is actually what a superb job the previous speakers have done of giving you a background to these issues. They saved me a great deal of trouble, and I think they've educated you a lot.

I think what I'd like to do is make a few general points, and then move on to really tell you more about the U.K. Biobank and the project itself, so you have a clear idea of what we're doing, how far we've gotten, and what it all might mean for things that you're considering as well.
So as Gil has already told you very well, U.K. Biobank is a project, it is not a single study. It is infrastructure. The aim is to support a whole range of studies, a range which we cannot really define now, in which we cannot define partly because they will be answering sheets of questions which we haven't yet phrased.

So it is a project to support a large number of studies with the overall objective of a better understanding of the way genes and environment work separately and together to influence health and illness. We are choosing to look at a large group. In our case, we define large as 500,000 participants.

I think what we've all agreed on is that the last decade of the last century saw biomedical science transformed by the Human Genome Project.

This is John Solstum from Cambridge. He had a role alongside many international colleagues in the Human Genome Project.

The Human Genome Project is truly staggering. But there is a danger that the project will become the museum exhibit of the 21st Century. I think it presents two challenges. There is a technical challenge. How do we take the human genome and work with that to produce science which is broader than simply sequencing the genome?

But there is also I think a moral and political challenge. How do we capitalize on that enormous breakthrough in science in terms of wider benefits to society and to public health in particular?

You could talk about going from the hype to the history. I think people will look back at this decade and say well, what did they do? They had the Human Genome Project, what did they do with it? The sort of things that they will look at are things like the HapMap, which I agree is an excellent project. We have to think, what else is there? We should be asking big questions now about what people will want in 10, 20, 30 years time.

Someone else said this is rather like planting the shade trees for the future. You have to think forward, particularly if you're talking about prospective studies. They take 10 to 20 years for the

real fruit to be borne. Because they have a long lead time, it in fact makes them very urgent. It means we must start them urgently. Otherwise, we'll have to wait even longer for the results.

But I also agree with David that there is a very important job to be done now. It is urgent, but we mustn't rush it. The detail work that we do now will determine the quality, the value, the comprehensiveness, and the scope of the results that people have in the future.
So what we have to do in the challenges to make sense of the data, we need to turn the data into information, and into knowledge. People like Sydney Brenner have come to epidemiology perhaps slightly late in his life, and has made this point very well. We need to start thinking not just about a genome, but about the distribution of genomes, distribution of genetic factors in the population, and what it really means for us all.

So to summarize, maybe in the 20th Century we had some discrete questions which we have answered I think very effectively. Things like the classic epidemiological questions of smoking, lung cancer, and other issues that perhaps we haven't tackled quite so clearly, and we have the genome sequences. We have very clear results from some of the biomedical sciences.

But we have to try and compile those together into meaningful 21st Century questions. I have just had a go, but one of them might be which HRT users will develop breast cancer and why, and you will have many others. I mean, as I said before, the questions are not known now, but they will arise.

I agree also with Gil, that many of these will relate to environment. Clearly nowadays we are much more interested in packs of smoking rather than individual smoking. We need to think. We need to be innovative. If we are merely contemporary now, then these prospective studies will be out of date. We have to think innovatively now in order to be contemporary in the future. Now, one of the things that you quickly get to when you start thinking about these questions is that the ideas of the size of current studies are too small, that you need very large studies. As Henry Ford said, "Quantity has a quality all of its own" in epidemiology, as in manufacturing.

This is part of a general trend in epidemiology and clinical trials. These are just some of the studies in the U.K. showing how many people were recruited, from 20,000 up to 120,000. So there is a general trend to recruit more and more people at baseline. In the U.K., we have the million women study which successfully recruited in fact, at one point, 2 million people. They overshot, they tried to stop at about 900,000, and ended up with 1.2 million.

So there are a number of things to learn from this. Firstly, there is nothing that we are trying to do with the U.K. Biobank that hasn't been done before by people in different studies, albeit on a smaller scale.

But the second thing is that these very big studies are feasible. They are difficult, they present challenges, but they are feasible. The public responds very well to them. I agree, again, with the previous speaker, that the public can identify with these problems, and the solutions to those problems. They know that we don't know all the answers, and they would like to help us to get the answer.

So what is Biobank? You've heard a quick sketch, and I'll try to just fill in a bit more detail, but perhaps take questions on further elements of detail later.

We are starting with 500,000 people. We have changed our age range. We have gone down to age 40 to 69 for reasons which I could explain. The essential idea is relatively simple. We

identified volunteers at baseline. We collect information on environmental exposures, we take certain measurements from them, they fill in a questionnaire, and then we take biological samples, blood and urine. We've considered various other samples, and we settled on blood and urine.

We then tracked those participants, taking advantage of the benefits of the U.K.'s National Health Service, corporation registration, and universal health care coverage, which gives us a very good start, but not all the data that we need. By no means all the data will come from these routine sources, but they are an extremely good screen from which to undertake additional validation exercises, including perhaps questionnaires in the future and recontact for validation.

I should perhaps say at this point by the way that we have taken the issue of environmental exposures very seriously. There is a subgroup set up on our Science Committee which is considering these. We have taken advice from the Health Protection Agency in the U.K., and environmental epidemiologists such as David Coggin are advising us on that.
The general point is that there is a lot of detail work going on on exactly how to measure exposures at baseline, which is being brought together by a number of subgroups advising our Science Committee. We plan to publish the results of that we hope by April of this year and invite comment, as we have done for all the other pieces of work that we've done. For example, the ethics and governments framework. So I hope that people in the United States will contribute to the process.

So here is the U.K. population in 2001. That's the U.K. Biobank corporation. You can see that the reason for choosing this age group is that there are broadly the same number of people in each age group here. This is the beginning of the slippery slope, I'm afraid, for most of us who were just in there. The major causes of death and morbidity start to kick in. I'm afraid from here on in, it is incidents of major disease outcomes. Of course, that's the point at which these studies start to be interesting.

There is an issue of how far back can you ascertain exposures. Some people argue, well, you really should be starting down here. You start with the children, because that's where the seeds of illness are sewn. We can debate the pros and cons of these. There is no answer to this. We need studies of children, and people are starting studies of children. We need studies of adults. We probably need studies of the elderly as well.

So it is important not to oversell these projects. Biobank is a big project, but it is only one part of a strategy to answer these questions.

It is a big study. There are lots of people in there who will develop lots of conditions, unfortunately. This is just to give you a flavor of the numbers. At baseline, within five years, we will have people with these sorts of numbers of conditions. So 8,000 people will have coronary heart disease. At the time, 7,000 will be diabetic, and 1.6 will have Parkinson's, and this is rheumatoid arthritis.

Now, these assumptions take advantage of what we know about volunteer bias. So quite a lot of work has gone into these estimates. We feel they are quite reliable. Importantly, there will be large numbers of people at baseline who suffer from various risk factors for disease as well. Therefore, we study the effect they have on people's health as they get older.
There are similar numbers for the numbers of people who would develop instant illness in the future. Gil talked about ten years. In fact, we plan to study people indefinitely. So we are talking now about 10, 20, 30 years. At 20 years, we will have 86,000 people who have developed

coronary heart disease who didn't have it at baseline. These are the sorts of numbers that you need if you're really going to get to grips with the interesting questions.

Scientific objectives. Very broad categories, but starting off with the public health aim which is to determine these separate and combined effects of genes and environment, and the nested case-control studies which you have heard about is really the selling point to the Biobank.

That was the one that really convinced the scientific peer reviewers that Biobank was worth doing. But nevertheless, you can also do cross-sectional prevalence studies, because there will be large numbers of people with diseases. If you choose the right diseases, for example, things like cirrhosis, you can do really rather nice studies on the cross-sectional studies on the prevalent cases, whereas with other conditions, you require instant cases.

We can also do cohort studies, the classic cohort studies looking at the particular exposure. Maybe an environmental exposure, or perhaps exposure to pesticides or some other condition, passive smoking, social class, or some occupational factor, and follow them up as a group.

An interesting variant on the exposure-based studies is genotype driven clinical investigation. We are recruiting a half million people, and there is every expectation that perhaps within five years it will be possible to genotype the whole cohort for at least a limited number of SNPs. It will then be possible to identify people with certain SNPs and invite them so they could volunteer in an appropriate fashion to take part in studies looking at the effect of those genotypes in the representative group of people, as opposed to people who you have identified because they are ill.

It is potentially very powerful. It raises a whole new set of ethical and legal problems even on top of the ones that Mylène described, I think. But nevertheless, we have had some quite interesting discussions with the relevant groups in the U.K. suggesting that this is likely to be feasible, provided it is done carefully.

The third big area of interest of course is in identifying biomarkers as early risk factors. Not just as a potential diagnostic tool, but it is something which helps us to explain the model, the fact that the substance is raised before someone has developed the disease may give clues to the disease mechanism.

In general, I think the point about this is that studies like Biobank and all the other studies we've talked about, and indeed comprehensive studies, will help us to understand disease models in a way that we never have done before. That of course is really the Holy grail of biomedical research. What we do with it is a separate question.

Particular scientific justification for prospective studies. Again, you've heard this before. Just perhaps one or two things. Having genetic information on people, regardless of severity, is important. If you take coronary heart disease, many of the people who develop coronary heart disease, it arises as sudden death. Not having samples beforehand can be a problem, or indeed risk factors beforehand.

Again, ascertaining blood samples, generally particularly for proteomics, not just for genetics, is very important. A general point about genetic studies is that if you take genes as just another risk factor, it is very important that, perhaps as Charles pointed out, you have to have no preconceptions about what the disease risk factor relationships might be.

If you start with case-control studies, you will very rarely detect relationships with diseases that you hadn't thought of. So if a particular gene causes Parkinson's rather than breast cancer, if you are doing a case-control study of breast cancer, you won't detect that relationship. So it's important to be able to pick up things which you weren't expecting.

It is important, finally, to be able to study health, as well as disease. I would argue that you can only really do that by taking samples of the whole population, not just a group of apparently representative cases and controls.

So to recap, the general benefits of U.K. Biobank lie in public health and looking at how these factors work together in populations, clinical medicine, understanding disease groups better, particularly looking at heterogeneity, 21st Century diagnosis, 21st Century prognosis as the essence of good clinical medicine, and bioscience. Particularly the biomarker disease associations.

The process of doing Biobank raises a whole lot of issues that we have had to work through. We think that will have some benefits for others, particularly our work on ethics and governments. The whole approach tends to provide better access to resources for scientists, and it promotes international collaboration. In some senses, it is efficient and economically beneficial as well.

Moving really onto the detail of Biobank itself. How is the U.K. Biobank funded? Well, these four research funders came together. The total cost of Biobank is 61 million pounds, about $110 million, of which the lion's share comes in the Medical Research Council and the Wellcome Trust, the Wellcome Trust being a large biomedical research charity, as well as the government, Department of Health, and Scottish Executive.

Is that a lot of money? It is approximately the cost of a Hollywood film. "Terminator 3" cost the same as Biobank. Some would argue that "Terminator 3" made a profit. Biobank may make a profit, too.

(Laughter.)

DR. NEWTON: Of course, the point there is that the value statement for Biobank is that the value of the resources is worth a lot more than the cost of collecting it. That becomes increasingly true as time goes on.

Another statistic, the health service in the U.K. spends the same amount in eight hours. So if we can have some benefit on health care, it will seem a small amount of money. Again, another comparative cost. The cost of Biobank is about 1 percent of that spent on biomedical research in the U.K. So funding a project like Biobank isn't really distorting funding priorities in the U.K.

That's my bit on the funding.

How have we established Biobank? Well, it is important to do this properly. It seems like very hard work, but I'm sure it has been worthwhile. We have a board, Biobank itself is a company, a charity with charitable aims, but an independent company.

There is a separate Science Committee which advises Biobank on all matters scientific. There is on the other side, a separate Ethics and Governance Council which is independent, chaired by a Professor of Bioethics which advises Biobank on ethics and governance, particularly in relation to

the interested participants. We'll continue to advise Biobank, and we'll speak publicly about whether Biobank is conforming to its ethics and governance policies.
In terms of implementation, we have six regional collaborating centers which represent scientific groups around the country, comprising 22 universities in all.

The general approach is to try to be as efficient as possible. This is a very large-scale process. If we're not efficient, we will fail. It is very easy to spend 61 million pounds and not deliver Biobank. I think it is possible to spend 61 million pounds and deliver Biobank.

It is an industrial scale process. I would emphasize the need for process and project planning early on. We've done a lot of that.

A distributed scientific collaboration is, I think, the only way to do this. But you do have to have strong central coordination. There is a potential to build a Tower of Babel in producing these big projects. There is a fine line to be cut between having masses and masses of talk and no action, and enough talk to make sure that you've covered all the bases you need to cover.

We particularly value the international collaborations. We've had a number of meetings with people in the United States which have all helped a lot. We do send out our material for comment quite widely. Again, we very much appreciate the comments that we receive.

So we will recruit participants. We recruit in the skill set from primary care, although in fact we are probably not going to use practices themselves that much. Essentially recruiting to the Biobank is rather like launching a new mobile phone. You've got to try to with direct mailing attract half a million people to in essence buy into your idea. So after considerable thought and planning, we are probably going to take more of that sort of line.

So we are going to start off relatively small and try and get the procedures absolutely right in the first year, and then roll it out in a mass way, taking into account this experience that you tend to overshoot in the end if you don't stop early.

How will participants enter Biobank? Well, they will attend the clinic. We have set up a dedicated clinic to do the data collection. Again, the efficiency of this process is so important that we think dedicated clinics are the only way to do it.

Samples are transported to a central resource, along with the data. The questions we hope will be on tox screen entry so that the data will instantly be amalgamated into the central resource as soon as the participants enter it. There's a big emphasis on archiving and curating the samples and the data for long-term use.

Of course, box number five is very important. It is always easy to forget this. In the end, the resource is only as good as the extent to which you can distribute and make available the data and the samples for future use. It is important to put resources into that now as well.

Data management is a big challenge. I'll just flip through this relatively quickly. We've got a lot of data acquired at recruitment to deal with the questionnaire, the samples, how the samples are stored, and the quality assurance data. At the end, we have information coming in from the NHS particularly, but also research input as well from dedicated follow-up procedures. The whole lot has to be amalgamated in a secure database.

There is also a lot of IT around the booking, scheduling, the managing of the process. All of this is new, and it has got to be developed. There is a lot of interest from the commercial suppliers, and we are working with some of them to develop these systems. Although mostly it is the experience of researchers that really tells you what is going to happen.

We also have a big investment in the U.K. in the National Program for IT. Many billions of pounds are being spent on drawing together these data sources, which may or may not be useful for us. We're not dependent upon them, but they would help.

Samples. Samples I mentioned earlier. We have done a lot of work on this. It was an expert group that pondered this, reviewed the literature, and produced a report which is available on the Web. We sent it out for peer review. In the end, we decided this is what we're going to do. We will get things rolling, but we think the mistakes we've made will be pardonable in the future because of the way we approached it.

In essence, we are collecting blood in various different ways so that they can be made available for the things that scientists want to do. Say there is going to be plasma and serum. We can do baseline hematology and baseline biochemistry. But the key to it is storing blood in such a way that people can do genetic, proteomic, and metabolic studies, as well as urine, particularly for metabolic studies. We also store blood, whole blood, so that we can immortalize white cells in the future, if necessary.

I just want to emphasize the volume of work involved, at peak we will be recruiting 750 people a day. That's some 3,750 bottles arriving in the lab every day. The storage will generate 24 million tubes, each of which are identified with two additional markers. This is a huge, huge resource, and it is quite a challenge to manage it.

The tubes we have stored in two ways. Traditional liquid nitrogen. You probably need that for whole blood in order to be able to immortalize white cells at that very low temperature. Putting blood into these things is fine. Getting them out is a lot more difficult. Traditionally, people have used liquid nitrogen storage facilities, and they are secure, so we will do that. But we also use an automated -80 storage.

This is a system where the tubes, you'll see in a moment, are stored in racks in here. These are held at -80 degrees. The robot operates at -20 degrees. This is a mock working factory, but it is very similar to the one that will be built in our storage facility.

The robot then essentially processes all the samples according to protocols, which are computerized. It uses a laser to recognize the tube markers. It knows exactly which tube it is handling all the time. They are extremely efficient. They are used quite widely in the pharmaceutical industry. They are used everywhere really, including restaurants who apparently have them for picking bottles of wine from their cellars. So if it is good enough for them, it is good enough for us.

Of course, the huge advantage is that you can set the thing running, according to the protocol that the scientist has defined. It can issue up to 4,000 samples a day, which can then be made available to research laboratories for analysis. Whereas to extract tubes by hand from liquid nitrogen, it can take up to two months to get 4,000 to 6,000 samples out. That's one person working for two months. It is extremely unpleasant work, if anyone has ever had the experience of doing it. There are health and safety issues.

So this is the way to go, this is the way to do things in the future. It is cost-effective on the sort of scale that we're doing. The cost of the -80 storage is about the same as the cost of the liquid nitrogen storage.

Ethics and governance. There is a huge amount that I could say about this. To summarize very briefly, Biobank is based on the fact that people are volunteers, and most important, that they can withdraw at any time. They give broad consent to future use, and this is a huge issue. I think I'd be more optimistic. I think broad consent has been quite widely accepted, particularly in Europe, as an essential approach to prospective research.

Now, the question of what broad consent means, and what safeguards you have to put in place to allow broad consent to be reasonable is a big issue, and needs careful consideration.
Data security and confidentiality have to be assured. There is a lot of work that has to be done on this. We have chosen to retain control of the samples. We think people are wary of their DNA being widely distributed, and therefore, we have tight control over the samples. But on the other hand, we have full access to evaluations and tests of the samples and the data for appropriate purposes.

Now, the word "appropriate" needs to be defined, so we have internal and external reviews of the science and ethics of potential uses at Biobank. One of the safeguards that covers a lot of this is our Independent Ethics and Governance Council, which volunteers -- we undertook a lot of public consultation before we started and drew this up. That was one of the issues that came out of that public consultation that people felt an independent group who could speak on their behalf was important.

We have also had a lot of support from Parliamentarians. We have done a lot of public affairs work with the Science and Technology Advisory Committees for the House of Lords, and for the House of Commons. In fact, there is a very big report from the House of Lords on genetic databases which was done I think as early as 2001, actually.

Biobank is a big study, 500,000, but it's not big enough, by no means. You quickly run out of individuals for a lot of studies. It is essential that we can collaborate. Collaboration means two things. It means encouraging people to set up similar studies and working with them, but it also means harmonization. It is no good if we all did studies which don't talk to each other, which is why the work at P3G is so important, and indeed the work of Muin Khoury's group from CDC, which looks at the other end of looking at the outcome of the research studies.

So there we are in the U.K. These population studies lend themselves to countries where you have population registration and universal health care coverage. So there is a natural tendency for countries like Canada, U.K., and the Scandinavian countries to think of setting up these studies. But as we've heard today, there is work going on in Japan, and there is work going on in Singapore. I was at a meeting in Sweden last week with a number of delegates from Singapore. We are very much hoping that the U.S. will make a contribution. Already there are studies such as the Marshfield study, which clearly will make a contribution. I would be astonished if the U.S. doesn't really make an important contribution to this worldwide collaboration.

Of course, you are very welcome to use our data. It would be great if we could swap.

How far have we gotten? Well, here is the timeline. We are starting pilot studies, we are doing some molecular pilot studies testing the sample handling procedures, and testing the clinical procedures. We'll start integrated pilot studies which will look very much like the real study in

September of this year. We start the main study in January, 2006. From then on, it is one person every five minutes for five years.

What are we doing at the moment? While we are looking so tired, it is very hard work. I have to say, it is very hard work setting up these big studies. There is a lot to do.

We are doing the piloting, we are setting up the IT infrastructure, and trying to design the clinical applications. The tox screen questionnaires are quite innovative. Very importantly, we are planning how we approach the general public, and developing a communications strategy to support recruitment.

The participants are fundamental to the studies. If you don't have the trust of the participants, if you don't convey the fact that we think that they are participants, not subjects, then people will walk away from us. So we take this very seriously.

We are developing this under the protocol. The protocol, which was published about two years ago, was really a proposal. There is a huge amount of detail work to be put into the protocol. For example, we mentioned environmental exposure measures. That in itself has produced a wonderful draft report, and there will be a second report. So there is a lot of scientific detail work to be done.

The Ethics and Governance framework will probably remain in draft throughout the project, because it needs to be brought up to date continually. We are thinking we will produce a new version quite soon. We put it out for public consultation. We are implementing the laboratory processes. We have commissioned our robots, and the people in Cambridge are building the robots. We are building the building.

This is where the automated storage facility is going to be. This is the new headquarters of Greater Manchester Police. This is in Manchester, U.K. So we thought this might be quite good in terms of putting burglars off, to be quite so close to them. These buildings will go up quite quickly. So we hope to have that ready by September of this year.

So what are the challenges? A number of challenges. Delivery against the timelines. It is a big super tanker of projects. It has got many, many people involved, some of whom have vested interests. It's important to try and draw these together behind a common goal.

The ethical approvals. We think we feel secure. We've had a lot of discussions. We think we have a lot of support. We have talked to all the right people. We have been absolutely straightforward about it, but it takes time. It is very difficult to bank on when you're going to get the final approval. So whilst you have your detailed project plan, the ethics committees can feature quite high in the risk management of that.

We need to negotiate access to all the information sources that we need, and we need to ensure continuity of the data chain over many years. By the time the people come to use the data, we'll all be long gone, so it needs to be carefully documented. Professionally, I should say, long gone. So finally, what is special about U.K. Biobank that perhaps marks it out? Well, certainly the size of the project. At the moment, I think it is the biggest funded project, both in terms of number of people, but also in the long-term nature of it.

The biological resource will be unprecedented. There was a great deal of interest just in the biomarker. So people would fund Biobank just to get hold of the blood samples. But Biobank is

a lot, lot more than that. The epidemiological design of Biobank is what really makes those blood samples valuable. Because the inferences that you draw from the analyses we think will be more reliable than inferences drawn from other biological resources.

We have, in terms of ethics and governance, an important element. We can recall the individuals, the participants, for intensive phenotyping, and for other information gathering exercises. So it is a continuing relationship with them. We are using written records extensively in the NHS, and we think that that will have quite wide benefits.

I think, again, to emphasize the ethical approach is one of public participation. We hope that by showing that this is an effective approach, that it will to some extent set new standards for this sort of work. Not just in the U.K., but internationally.

Thank you very much.

DR. TUCKSON: Thank you very much.

Kevin, you had one quick question? We'll just do this one, and then we'll go to the next panel.

DR. FITZGERALD: Yes, thank you.

Just a quick question. You keep talking about the public participation, and the participants, not subjects. Do you have outlined a process for how these participants will participate in the process?

DR. NEWTON: In terms of influencing decision-making and the managing of the project?

DR. FITZGERALD: Right.

DR. NEWTON: Well, we have a participants panel, and we have been consulting with them in general.

DR. FITZGERALD: Okay.

DR. NEWTON: We have representatives of the public on our Ethics and Governance Council. What we've avoided is a sort of token member of the public on the board, for example. So I think we're open to ideas, particularly from our panel about that.

DR. FITZGERALD: Thank you.

DR. TUCKSON: Thank you so much, John. I appreciate it.
.