I'm David Bentley, and I'm currently Vice President and Chief Scientist at Illumina.

I was born in Windsor, England in 1958 under the eaves of the castle.

My parents -- at the time my father was working the theater. He was a musical director at the theater Royal Windsor and my mother was a biology teacher.

There was. My mother started it off because she was actually in Cambridge at the time of certainly the protein, structural protein work that was going on. And she knew some of the people involved. Dorothy Hodgkin and so on and she had knew of many of the people who were really developing that field. So she started me off. She knew some of the people who were writing popular articles that I was reading. But I'd also single out a biology teacher in high school. My secondary school which ironically called Watson. No known relation but Ian Watson was a tremendous, passionate, enthusiastic man and he had the benefit of almost a whole year off curriculum. We were able to be taught whatever he felt like talking us -- teaching us. So he taught us a lot about molecular biology. And also about the whole convergence of genetic inheritance with the molecular basis of it and the chromosomal basis of inheritance. It was clearly one of his pet topics and I just lapped it up, I loved it.

So, there are many people -- to many to tell or even remember but Cambridge was a wonderful time for me. I read natural sciences which really provides a blend including chemistry, in particular, and my link to the chemistry department of course became very important for later on. But really, one of the tremendous tutors there, David Hanky, who was in my college, Jesus College, really was a tutor to me throughout the time. It was a lucky break. He was teaching me for much of the first two years. And that one to one or one to two tutorial was a feature of the Cambridge system that really allowed us to explore areas in this case of biology and to really pursue our passions together. He was a tremendous guiding light in the -- then I went on to biochemistry. That was my final year subject and a tremendous course which really gave me a broad view of biochemistry without too much specialization I think, genetics wasn't a particularly strong feature of that course. But we'd had a lot of genetics already, so that was really the basis for my degree. It was a three-year degree. Really rather short by today's standards and I certainly, by the age of 21 I was considering my next step.

So, my first PhD year was in the laboratory of molecular biology. I was in Fred Sanger's division, protein nucleic acid chemistry. Wonderful time to be there, 1979, 1980, and it was just at the time when the whole concept -- first of all, being in an environment that simply pursued technology. They really were less interested in the biological problems. George Brownlee had been Fred's first PhD student I think and so I was one of George's PhD students so continuing the lineage. And I had the pleasure both of working with George directly and George was a great innovator. He liked to do different -- he liked to do something no one else was doing. He'd spent a lot of time working on RNA sequencing in his earlier days and that innovative and adventurous spirit showed through in this remarkably quiet but enthusiastic man. And his enthusiasm just came through. And so he encouraged me to do something very different as well and tackled the -- probably in hindsight -- pretty impossible problem to solve. And I was left to explore and to learn

from many, any of the people in the laboratory of molecular biology which is a tremendous environment for me.

Fred is a remarkable man, was a remarkable man. He's quiet. A quiet presence in the LMB. Great to see him strolling down the corridor every morning to get his ice from the ice bucket, ice machine and go back to his lab. And he was also the holder of the all-important P32, the P32 DHEP which we used for all the sequencing. And you had to go to Fred's lab to collect your [unintelligible] of P32. And certainly on occasion I would see him standing there. There's a photograph in fact of him which is just like how I remember him, standing there with his red cardigan on staring at the [unintelligible] graph that to his apparent surprise it really worked rather well. And Fred was a really modest man but of course at the same time very quiet, very thoughtful, and very direct actually in the way he talked and it was a pleasure to get to know Fred, to talk to Fred. And I had a number of connections with Fred later on of course because later on Fred actually did come to the Sanger Centre. He blessed it, allowed us to use his name and even more recently George Brownlee actually was writing his biography. So small world. We came together again and George actually asked for some help with a chapter in the book. And then of course that all came to a head when Fred died so the biography really became very timely and was published shortly after. So it was wonderful really to still be there and to have a change to perhaps acknowledge and reflect on Fred's enormous contributions in a very quiet way, in a personal way as well as of course a scientific achievement that Fred really brought to the whole field. He was a great mentor and guide as well as tremendous scientist.

Yes, after a year George Brownlee was already actually about to leave the LMB and set up at the Dunn School of Pathology in Oxford so my PhD got transferred to an Oxford [unintelligible]. And four of us including George was a small nuclear group that moved across from the LNB to Oxford and we set up the lab and that was a remarkable experience. Very enjoyable. We really had to do everything including borrowing equipment and driving it over to Oxford to set up the lab and the LNB were very helpful and supportive of George and setting up. So we had a chance to really just set up a lab from empty to doing the first Sanger Dideoxy sequences in the Oxford lab some two months later which was a sign of success of the transfer that we were able to transfer what at the time was a fairly sensitive protocol -- fairly complex protocol. But it was a great time, yeah. That was actually the same time that Fred Sanger was awarded his Nobel Prize and it was a very curious moment. I was actually driving back from Oxford to Cambridge on the day. And I heard the news on the radio. Fred had just been awarded a Nobel Prize for sequencing so this was along with Maxam-Gilbert as well. And I got to the LMB and I couldn't believe it because the lab was completely empty. I almost thought it's not a weekend, what's happening here? The lab was completely empty. Nobody was in any of the floors of the laboratory. I thought this was strange. And of course, sure enough, they were all crammed into the canteen up on the fourth floor and rumor has it that [unintelligible] two others had bought up the entire champagne supply from Cambridge and it was being busily used to celebrate Fred's second Nobel Prize. It was a wonderful moment.

So the work in the Dunn school of pathology certainly went back a long way but I think covered many areas of cell biology in particular Henry Harris the director at the time had put a great contribution of cell biology to it. And so in that sense there was a breadth of research going on in

the Dunn School. And like many things, I think in molecular biology you come along with a technology and the question is how it will have an impact on the environment around you. George in particular interacted not just with the Dunn School but with the other departments close by as well. Biochemistry in particular. And set up collaborations where we were -- we had the chance to see how the molecular biology, the techniques, so cloning as well as sequencing of course was a predominant technology at the time and it had enormous universal applicability to many research problems so there were great collaborations being set up. One of the very prominent partnerships which George formed was actually with the then professor of biochemistry, it was Rodney Porter. And who—cause he ran a tremendous protein chemistry lab and so there was a tremendous discipline and trying to understand proteins and characterize proteins through purifying them and characterizing them at the peptide level had been all the rage. But it was getting more and more difficult because the proteins that remained to be discovered of course were in much smaller amounts, vanishingly small amounts difficult to purify, difficult to know if you'd purified the right thing. And so to the protein chemistry lab the excitement of extending the characterization of the proteins involved in immunochemistry -- which is Rodney Porter's particular interest, the compliment proteins -- suddenly became a ripe target for molecular biology to lend a hand and to start to find another ways of searching through the nucleic acid based approach to find these elusive genes or messenger RNAs. And that spawned the whole field in Oxford of eukaryotic molecular genetics, human molecular genetics through George's lab. George then attracted a number of visitors who were very influential, particularly people who were involved in hemophilia, hemophilia b. And so George embarked on a pretty extensive program to clone the Factor 9 gene and that was a successful approach and others after it. And on the back of that experience, that pioneering experience of how to get a gene cloned from a little bit of protein information. It was something which was replicated time and time again and so that was the basis for much of the characterization of the compliment proteins a few years after that. Or perhaps only a year after that. And so I was working then. I transferred then from George's lab at the end of my [unintelligible] to work directly with Rodney Porter and to continue that gradual dissemination of the molecular biology techniques from one department to another. To the biochemistry department and the Dunn's School tremendous access of collaboration in Oxford, something which I enjoyed for a number of years.

So I guess I should put a couple of things together. I'll say one more thing about George and Rod in the contrast which is a fascinating one. Because I mentioned already George was very much innovative, tried to do something different. Rodney actually was almost on the other -- on the other extreme. Rodney believed in doing the obvious. If there was a job to do you should get it done and you should not hesitate. You should not try to think of the less obvious experiment but you just march over the ground and characterize things. And we did. We hit a seam with molecular genetics, human genes, and we did a great deal. It taught a great deal about the productivity that you could actually engender by developing a field and really working with it and expanding the applications and collaborating more and more widely. And that was an interesting contrast in the style of work to George. Both are incredibly valuable, incredibly valuable training. It was still in George's lab both before and after I had moved. Not only some of the key people involved in hemophilia and one or two other medical, genetic subjects but in particular there were two people who both joined George on sabbatical for a year in George's lab. One was Ted Freedman from UCSD who was very interested in gene therapy and he'd known George for a long time I think.

And Ted was a wonderful mentor, really great guy to have around and really took some time with us to teach us something of what he thought about where things were going. The other was Francesco Giannelli, a hematologist from Guy's Hospital in London and he was my next boss though I didn't know it at the time. And he in particular, both of them actually collaborated very closely with the lady who then became my wife. So I met my wife in the same laboratory as well. And my wife was also working on hemophilia B. And so together there were projects that increasingly became relevant to patients and the move from the appetite to find the gene through knowledge of the protein moved to now we have the gene. Now we can really get access to the genetics. And of course, the genetics really involved this is a means to identify mutations and conditions and that's where really George and Francesco in particular with others. Charles Ritzer in Oxford and set up the suggestion they would collect DNA samples from hemophilia patients and actually characterize them at the DNA level to search for mutations which were involved in the cause of hemophilia B. And that was really for me, that began to -- it began to open my eyes to some of this, the medical aspects of it. Because Francesco was a hematologist he worked in a pediatric research unit in Guy's Hospital. Certainly when I got talking to him rather more it became rather more obvious that this was a whole new direction of research to go in.

Well, yeah, the human genome project -- not by name for three or four years more I think. I guess I heard about it when I'd already moved to Guy's. 1986 was a very important year for me when really I began to hear more discussions about the human genome project. But the concept of mapping and characterizing human chromosomes was actually rather earlier. And in looking back just recently I've reckoned a few times there was a project way back during my first post doc with Rod Porter where we actually mapped four of the compliment class three genes together in a small contig but at the time it was a huge effort to assemble a small contig of four cosmids that overlapped, that were shown to overlap and they identified this cluster of four genes. And that was the nature of physically mapping out in cloned DNA to try to replicate or reproduce the pattern of the genes as they actually sat in the chromosome. And this for me was really -- it was something you could almost feel physically. It was actually characterizing something that you could definitely prove to be right. Interestingly it was never the complete picture. There was always more to discover. But that was my first contribution to rather rapid coming together of a contig of clones to map genes to begin to understand something which really nobody had had a concept of distance along a chromosome before. And from that -- this is where Ted Freedman and Francesco Giannelli and George all come in. And again, they started from sitting on the Factor 9 gene and actually the Factor 8 gene next to it which George was involved in for a while, to actually again, asking the question well, could we link up genes on the X chromosome? Of course the X chromosome was particularly an exciting chromosome because of all the diseases that were known to be associated with the X chromosome. The homozygous nature of the male really immediately manifested -- in males recessive conditions immediately manifested themselves. So hemophilia B and hemophilia A were only two of many X-linked diseases. So, suddenly the idea of taking this whole concept of mapping rather further linking genes up and then perhaps being able to search in the material in between and to really begin to capitulate the linear nature of DNA and its ability to code along an entire chromosome and to start a map back the concept of linkage at the molecular level. This was my exciting year at school all over again. The ability to start characterizing the unknown marching through it. But Rodney Porter had taught me don't think too hard about it, just do it. And that was an exciting moment. So that was really where the idea of mapping genes and looking, really taking advantage of the linear continuity of DNA and walking along the

chromosomes in some fashion took root. So this I think was somewhat before the human genome project was properly defined but nevertheless, the concept was there and I'm sure many people around the world were having experiences like this. And so when the concept of the human genome project was vocalized, formalized it made a lot of sense.

The geneticists were branding people with particular qualities which I'm sure that it's not really true. But if an individual let's say has been working on the genetics of a particular disease which involves a particular gene then that's one thing and they rapidly go down the medical route of understanding, seeking to characterize the genetics of the disease. But when you broaden the field to applying the concept to what you know a lot about in terms of aging and a disease and the mutations that may cause it you recognize instantly that that can be applied to any genetic disease and probably many we don't even know are genetic. It becomes a universal principle. And that's where genomics actually helps. So genomics is not there to steal the genetics from the geneticists. Genomics is there to really help and support and make genetics much more accessible to many more diseases, many more patients, and to provide a much more streamlined process for characterizing the molecular genetics of disease. So I wouldn't claim to be doing genomes before the human genome project. The human genome project was a wonderful description of the concepts, the isolated events, the examples which I had seen and suddenly it came together. Another very important element in all this which also -- on the one hand there's the human genetics that makes the interest and the utility of the human genome sequence so important. The promise of the human genome project is that it will help and revolutionize genetics and medicine. But it was also the work on other genomes that was more advanced that wasn't motivated by human genetics but it was nevertheless motivated by the same idea of characterizing a complete organism at the genetic level and here particularly, and I will pick out the nematode worm because that's one. I didn't really know John from when I was at the LMB but I did get to know John when I -- shortly after I arrived in London. And I was just beginning to absorb the genetics of X-linked disease at Guy's Hospital and become immersed in that. But I was also there actually on purpose to do research on areas that could not be funded from other sources. That was the terms of the generation trust, a private trust fund which was funding my research and so that's where chromosome mapping was really one of the things we tried to get our teeth into back in 1985 I guess this was. And I went along to a London molecular biology club seminar and there were three talks at it. Bob [unintelligible], John Sulston, and Peter Little. And they all talked about the same thing in the sense that they talked about establishing contigs of clones that represented different organisms. John of course in particular was talking about the worm genome and their work and beginning to assemble contigs of clones to characterize the complete genome. And I'm sure I don't know this, but I anticipate anyway, that John's interest was always to characterize everything about a living system. John heard previously and he told me on more than one occasion the happiest time of his life was sitting in a tiny room with a microscope at one end actually staring all day and mapping out the lineage of cells in the nematode. And of course he eventually, I believe established the complete lineage of cells in the nematode as well as discovering phenomena like [unintelligible] and programmed cell death as it was originally I think coined, the term is coined. But again, the idea that we -- I've done that, so the next thing I do I'm also going to have an appetite to try to cover the whole thing. And that was clearly behind John's motivation and I loved that. The idea of doing the whole thing, getting the whole job done. Doesn't matter how long it takes, maybe a lot of marching but march through and do it and don't stop until you've done it. And one of the other important influences is Maynard. Maynard Olson. I met Maynard again

fairly early on in my time at Guy's. And there are two or three threads where I met Maynard but this particular one is when he came to Guy's hospital to a course, a Wellcome trust advanced genomics? No, advanced genetics course. I can't remember the name of it now. Wellcome trust advanced courses anyway. This particular one three of us were teaching genomics, the technology to various people that were very keen to start. Maynard was one of the keynote seminar speakers. And after the talk we ended up drinking beer in Guy's Hospital somewhere in one of the older parts of the hospital. And he said in his view that the yeast mapping project which had been going on in parallel to the nematode-mapping project -- Maynard and John knew each other very well -- he regarded it as a failure. And I said, why on earth do you think it's a failure? Maynard said, well, there are six gaps or however many there were. The idea was to get continuity and I couldn't close the gaps. I hadn't closed the gaps. He was missing stuff. And it was a bit like, it was very like the approach that John Sulston was taking to looking at the whole problem and until you've done it you haven't succeeded. But Maynard was rather more purist about it. And I said to Maynard at the time, well, I don't agree with you. Maybe the job isn't finished in terms of continuity but look at all the stuff you have done. Look at the value that's in the 99 or so percent that you've got. And I think I said at that moment, if I had a human genome that had a few gaps in it I'd still be pretty happy with the outcome. And Maynard thought about it and I don't think he necessarily commented on that much. But, clearly it was an interesting concept and a very good one to try to motivate to get complete continuity not to stop. And of course, these genome -- the worm genome and the human genome in particular -- still has gaps. There's no question. And those gaps become a point of debate both of the ones in the achromatic sequence, the odd percent, what's one percent? Well, actually, that's 100th of all the genes maybe. And then of course, we don't necessarily usually refer to the heterochromatic regions and those things which are almost completely uncharacterized. So, completeness is certainly a relative term and so the concept of Maynard and John striving hard to really get the job done and completed was an important one and probably drove both my understanding of just how determined you have to be, how much you have to slog through and how much you have to be looking very hard at methods that seek to achieve the goal but don't actually look encouraging when you look at them close up. There are methods that were published that really were claiming long range continuity but there were shortcomings and you have to look very hard and critically -- self-critically often to actually try to adhere to the concept of doing a really good job. And ultimately faithfully representing a piece of DNA in its entirety in some immortalized form whether cloned, biologically or ultimately of course sequenced and stored as information.

Maynard's a great thinker and a great communicator and he thinks both very clearly not too much in the detail but he will certainly try to see right through a problem from basic principles and I think try to create something I think very individually, create something almost from scratch. He will not spend too much time basing his own ideas on other people's theories necessarily. He really will certainly question other background information, other theories and really build something from scratch and I think there's a purity in there and a clarity which is tremendous. He's also a great communicator because he will evolve a whole theory, strategy, whatever almost without pause. You can sit and listen for 20 minutes and you learn a huge amount and he never really deviates from the subject in hand. And it's clear that he just has absorbed a huge amount of background and perspective to enable him to see clearly and to concentrate on the important things in his mind and I think I agree with pretty much everything he thought about. And it's a tremendous synergy that I felt or respect for his approach. Because I fell into his way of thinking

quite easily. It was easy to follow the thread. There was a strong thread there, there was logic and it was -- it was I say very easy to follow Maynard's thinking. Also great philosopher particularly on the science. If you've seen Maynard quite often I think gets given the task of summing up an entire conference. And he goes around getting opinions from everybody and just in three days he not only manages to absorb all this new information that's coming out at a conference in a very concentrated fashion but to somehow distill it down into some very clear messages and some very strong messages and to communicate them at the end of the conference. And that's a remarkable capability. He's also very hard to put off. Because on one occasion when I think he was doing the summing up, I think it was a conference that Eric and I were organizing, but actually the lights went out in the middle of Maynard's summing up. And apart from a quick quip and a pausing he simply carried on. And the clarity and the drive to get to the end of his point was absolutely crystal clear.

In the years just coming up to the Sanger Centre I -- I was following John Sulston's work quite closely on the nematode worm. I'm very attracted by the global view of trying to do a complete job on the nematode work molecular level. The mapping level this was, this was before sequencing. And I did actually meet John a few times and shared what I was try8ing to do with human DNA. And I was trying to really apply a lot of the technology and lessons learned from the worm to mapping human chromosomes and we went further. We actually collaborated to start using C. elegans software tools, HTB in particular is when I met Richard Durbin. And so quite along contacts were formed with John and John was very aware of what I was working on. Sometime later he actually -- when I actually published a paper on a very small piece of the X chromosome that I managed to map both in yaks and in cosmids done by a student of mine, Jill Holland. And he actually confessed after that he was very surprised when we got it to work. He didn't think we'd get it to work because of all the repeats in the human genome that prevented this from hybridizing from one level of the genome to one level of the reagents to another. And I don't know how much that counted but we did manage to get a great deal to work. We used his finger printing techniques. We used his filter based hybridization strategy and employed it all on human chromosomes. On the X and on 22. And clearly we'd been working quite closely with Richard Durbin. I can't quite remember when it was 1991, maybe 1990. And as it turns out this was sort of about the time when the whole -- the beginnings, the early foundations, or perhaps the seeds of the Sanger Centre were sown and the idea that both John and Bob -- Bob [unintelligible] needed to be given a foundation, a more secure foundation to work on the worm and on the sequencing of the worm genome. And so, there was a remarkable moment when actually -- we actually, we were hosting a visitor in Guy's Hospital at the time in my department, Kay Davies who was being hosted by Martin Bobrown, the head of the department. And happened to be in Francesco Giannelli's office where I was as well and so we were talking and the phone went. And Francesco picked up the phone and then passed on the call to me because it was for me. And it was John Sulston. And I wasn't really expecting a call from john. And John said, "Oh hello, David. How are you? You're well? Yes?" And he said, "I'll come straight to the point." He said, "How would you like to come to Cambridge and join me in setting up an institute?" And I froze because I felt that well, Kay and Martin and Francesco surely had heard the entire conversation going on. But it was a remarkable phone call and we realized it was another time would be better to talk. But that was the beginning. Certainly of course, when something like that happens you suddenly feel the whole landscape change. And of course the mind starts to work very fast on what the opportunity might be. It just changes all the conceptions, the previous conceptions and previous thoughts. And a very

stimulating time. I was very happy at Guy's. I was doing a great deal with human genetics at this point. But clearly this was another opportunity. Perhaps a difficult decision but clearly as time went on, a relatively short space of time, it was clear this was a very big opportunity. And an opportunity to really get much more involved in a very new venture which really drew together many of the early experiences that I had. And so I think it was less than a year later the Sanger Centre grant was awarded and I moved some six, nine months after that to set up a group and to bring some people with me from Guy's who had been working on the X chromosome and chromosome 22. And that was the basis for the human genome component of the Sanger program. So that's how we really got into the project. And I should say as much of the pilot studies we'd done at Guy's were funded by the MRC who set up an HGMP, a ring fenced program for funding small projects. But it was clear that if this job was going to be done, it needed a whole different scale of investment. And I think both the MRC and the Wellcome Trust both worked together to make that happen and to create the Sanger Centre. So at that point I think I was at least involved in genome projects. When the Sanger was first set up it was actually set up to support John to sequence the worm and Bob [unintelligible] in fact to sequence yeast. And I think it was more a question and to work out how to tackle the human genome. What to do, how to contribute to the human genome. And it was, I think almost every year after that there were ongoing discussions with the Wellcome Trust and proposals to continue and extend the program and to really start working on the human genome. And I guess it was probably about a year after that that we went for additional funding. And indeed, the Wellcome trust agreed that it was time to do more directly -- more direct work on actually working out how to and actually starting on sequencing the human genome. And that was when chromosome 22 and the X chromosome had been going on for several years before it and they'd come up from Guy's. Became the center of the human genome program along with a piece of chromosome four if I recall which had the [unintelligible] disease lockers in it somewhere though it hadn't been found before.

I think there was no limitation in will. There was very definitely a will to do it, so I think the two major limitations were probably -- it was resource-limited. It was funding-limited, although the Wellcome Trust are being very generous, and I think there was still a general sense of caution, that actually it wasn't unanimously necessarily the right thing to do, or to do it this way, or was it going to cost too much. I think there was still some debate about should we be waiting for new technology, or perhaps the debate had happened and we'd moved on, but of course not everybody had necessarily really satisfied themselves that the technology was ready, and I think there were technology gaps. Clearly, sequencing was the technology that was chosen, and that lasted, stood the test of time for entire human genome, but there were gaps. I think YACs kind of came and went. They weren't sufficiently stable to actually contain a faithful copy of the DNA within them, although they were quite good for a long-range continuity. Cosmids were too small; perhaps some level of bias. So one of the important gaps that was then filled was the PAC and the BAC system. That was missing at the beginning of the program and that was an important transition, filled a really important gap. I think one of the other technology gaps was actually a framework to independently verify, demonstrate, and even create the higher level order of contigs along a chromosome, and this where both first [unintelligible] genetic map, and certainly in the U.K., the eyes were on the European initiative in France in the [unintelligible] laboratories in Généthon [spelled phonetically] where Jean Weissenbach, of course, created a microsatellite based map, very elegant piece of work that really provided a level of continuity, albeit the points were somewhat far apart, followed relatively quickly by the radiation hybrid mapping, which was a very exciting

time indeed, where we actually went from being -- having a system that could not only make use of non-polymorphic markers, and many more of them, and achieve higher densities, but they also could integrate the Généthon markers as well because it was all PCR-based. It was all STSs, a concept that Maynard and his colleagues had advanced, and suddenly automating very high throughput PCR reactions to map out a set of markers much higher density than the genetic map. That really provided the density of markers needed along a chromosome to allow the clones to then take over, and we no longer needed YACs to provide the continuity; the distances could be filled with PACs and BACs. And of course John Sulston's fingerprinting method which still really enables us to link clones together. So we now suddenly had two approaches, orthogonal approaches, to create the continuity that was so important. There was the fingerprinting of the PACs and the BACs to provide the continuity, but there was also the reference points from all the STSs anchor points to actually identify individual clones underneath those markers, and if we needed more markers we could now do them. We could generate more markers, do more PCRs, and actually add markers to the same map, and that integration, that very close integration between the genetic, the radiation hybrid, and the clone maps, was a very important area, two technology gaps filled in fairly quick succession by the pioneering efforts of laboratories around the world. Radiation hybrid map famously brought together the David Cox lab and Peter Goodfellow's background, and of course the concept actually came, in part at least, from Henry Harris in the Dunn School of Pathology some years earlier in constructing these mouse-human hybrids that became the reagents that underpinned the radiation hybrid map. So those technology gaps were important to fill; they led to very constructive international collaborations. We were exploring how to collaborate, meeting the whole community through these efforts, and that, I think, quite quickly really stimulated the funding agencies, probably, to do more and to consider that there was really -- there was a coherent approach that was very step by step. The Human Genome Project must be the most step-by-step hierarchical study of any genome, I think, that's ever been done, but that was partly because we were working every step out at -- one after the other, rather sequentially. We were relying on different techniques to try to tackle a very large problem, and also I think the community in general, and perhaps the funding agencies in particular, needed to see those levels of evidence, the levels of proof, the levels of being able to obtain completion at different levels of resolution to be confident of moving on to the next level, and I think that was a fairly compelling element of the project that enabled us quite quickly to move through the barriers, and to confidently expand the program.

Clearly, the sequencing technology in multiple labs and within the Sanger certainly -- John himself and Jane Rogers took a very personal interest in putting the technology providers through their paces. But the sequencing was really coming on apace, I think, although it looked a little shaky to start with, the automation of the -- to get fluorescent detection instead of radioactivity. I think some of the early limited cosmids -- there was at least one person who was convinced that the right way to do it was [unintelligible] radioactive sequencing while the first fluorescent machines were not really delivering so much, but it was the right thing to do, no question. Embracing a new technology; getting away from radioactivity; a high level of automation. They were absolute bellwethers for progress.

It's interesting. I think from the very beginning the sense that the human genome could be done, from my perspective, and it was based very much on the argument, well, one, there's 100 megabases and that looks like it's going to get done. In fact, it was probably largely done by the

time we came to this point, and done to highly automated sequencing, but also finishing and all the hallmarks of a high quality product were there. And each human chromosome is about the size of a nematode genome, so all we have to do is divide the project up by 20 or so and we have 20 nematode projects, and that seems to be really quite manageable. That's one example of the philosophy of going chromosome by chromosome. But I think it worked quite naturally, because that's what people had been doing for some time. We'd been working on 22 and part of the X chromosome at Guy's Hospital, and that illustrated two things. The reason that we chose chromosome 22 was actually to get away from a busy chromosome where lots of people were working on different parts, because -- fine! We would all discuss and work out, and some people were working on the same bits as others, and we'd figure that out at some point. I think Eric Green and I actually had a go at suggesting as long as you went -- stuck within your own gap between two STSs, then we would actually carve up the entire genome STSs by STSs. That never quite worked, and instead, chromosomes were the much more logical currency. So we moved to chromosome 22 because it was essentially almost unstudied, at genome level anyway. It was also small, so continuity and finishing the map was going to be much quicker than the X or an average size chromosome, and at sum-33 megabases, of course, was only a third the size of the worm genome, so things were starting to look up, that we could actually make progress quite quickly. So between those two examples, chromosome 22, we were simply looking at the whole thing. The X chromosome, bigger chromosome, popular chromosome, brought with it its problems both in terms of size and the complexity of the community. So those two models rather played out for the whole genome, I think, and others were doing -- I think following probably exactly the same logic. David Cox and Rick Myers were interested in chromosome four because of an early interest in certain genes on it, I think, chromosome six, and others. Gradually the splitting-up of the genome organizationally into chromosomes reflecting the natural interests of different laboratories was how I think it really took place. And I do recall one or two meetings where we got together quite quickly at the beginning of one coordination meeting to try to divvy up the chromosomes between the groups, and it kind of worked quite well because most groups were actually working on different chromosomes. There were one or two debates, and the X chromosome was always going to be more of a mosaic, but some of the others were more straightforward, and alliances were formed. People began to realize if they were all part of something bigger, and actually you had a chance to look at the rise [spelled phonetically] for the whole genome, then it made much more sense to try and get the whole job done than to fight over one chromosome. So I think the idea of splitting teams up between chromosomes worked very well, and worked at a particular meeting. It was a very different agreement to work on certain chromosomes, and this was about -- I'm not sure, '95 or '96. I would say '95, but I'd have to check on that. It coincided with the Wellcome Trust agreeing to fund the Sanger for more, and so that's when we started chromosome six. Within a week we had a chromosome six team. It was astonishing. We took people from 22 and from the X team and formed the chromosome six team. We suddenly were able to expand onto another chromosome and clearly illustrated the scalability of the whole program at the human level as well as at the technology level. One of the great things that happened at that time -- Wellcome Trust had funded us for a sixth of the genome, and the other members of the G5 and other labs as well were funded to varying degrees for other chromosomes, and I once drew up a poster which summarized -- I boxed each of the chromosomes in colors to represent each of the people contributing. And about a third of the chromosome were not assigned. We had a royal visit at Sanger. Princess Anne came to visit the Sanger, and I think officially opened a building, and as she walked down the corridor, and I had this poster up on the wall, and she actually stopped to say

hello. Princess Anne is not a geneticist. She's not a biologist, but like other members of the royal family, they're very perceptive and knew the questions to ask. So she looked at this diagram and she said, "Who's going to take responsibility for the other ones?" And that was such a good question to ask, because of course -- it was exactly the right question, and of course it was eventually the question that got resolved maybe a year later, or less than a year later, when suddenly everything scaled up to the concept of "we need the whole genome; we need to do it now; we need to have a strategy. What's the strategy? We need to make sure there aren't any parts of the genome that aren't covered." And then I think the chromosome by chromosome strategy really did expand. More resources became available as Sanger funded -- was funded for now up to a third of the genome, and at the same time, I think, the NIH must have done a lot to really stimulate that. The DOE, I think, came in possibly more firmly, so everything happened at about that time.

I think it's important to say that the G5 didn't do the entire genome. There were contributions small and large, and I think that was a very important concept. I think the simple solution to it is that if you release your data and you share your data in some standardized, coordinated fashion, then it is clear that you are working and contributing to the whole, and it makes no sense to duplicate it. So, in that sense, a small group can survive under certain precepts like being coordinated and being transparent and showing and sharing their contributions. Clearly, the question of scale and being able to make economics of scale is very important, and that's inevitably a driving force. This may have been a decent level of automation around the sequencing, but there was still a huge amount of potential economy of scale as a result of really building dedicated teams and dedicated laboratories that really did this. Sanger was like that. Sanger was working on two and sometimes three shifts a day. That was not something a small lab was going to necessarily keep up, so inevitably there was a question of cost overall. But I think it's -- I'm delighted the smaller labs, the labs that contributed smaller parts of the sequence, did stay and contribute because they added a great deal, I think, in other ways to the project. And of course one of the latest and most recent members of the consortium, China, came in late to the game, but contributed some sequence to the program, and of course the outcome of China becoming part of the Human Genome Project of course had huge ramifications for the future in a very global way. I think, clearly, if that had not become possible and if they hadn't been welcomed to the community, then the world and the Chinese sequencing community would be the poorer for it, or would possibly not have evolved.

One of the very interesting things that the community tested out somewhat before the G5, back in the first Bermuda meeting in '96, was the decision to create this framework of data sharing, transparency, and to begin to set standards about what was being generated. But the data release policy demanded that the requirements of the project were put ahead of the constraints of individual nations or laboratories, and that was a hard decision for some people to take. I think one or two people actually couldn't, that they had to go back to their national constraints, their governments, and decide whether they could participate or not. But achieving that transparency, that sharing, really formed the basis for setting of a common standard, and that's when both large and small groups could participate as long as the standards were set and the necessary constraints were met.

I think people were always talking about SNPs to some extent during and before the program, really. SNPs, any type of variation certainly, whether it was an RFLP or not, was very important

in genetics. In that sense, it was all being talked about. I think the concept of scaling up clearly became possible -- I think it was starting to get talked about before the genome was assembled. That's certainly the case. There was a fascinating transition once the concept had been established that it was a really good idea to develop a SNP map of the genome, or to collect a large number of SNPs for the genome. I guess it embraced the idea that here we were being able to look genome-wide. Here we were -- the concept, perhaps, of the STS was really, again, providing some momentum to the idea of being able to spread markers right across a genome and the power of doing so at an ever higher density, plus the fact that there were ways of doing it through sequencing, and sequencing was now much easier to do at scale, and suddenly that was not the limitation. Only six, seven years earlier, sequencing had been an impossible, cost-limited element to doing an experiment, and suddenly sequencing was a currency you could work with and generate large amounts of sequence data from other individuals, compare them to each other and later on to the genome to actually identify variations systematically at speed. Fascinating transitions happened during the SNP consortium, and we're skipping over the SNP consortium a little bit here into the early part of experimental strategy, but when we first started the SNP consortium, the centers that contributed to it, we started an approach that did not assume the existence of the genome sequence. We started on a process where experimentally we targeted specific subsets of the genome in a genome-wide fashion, so we would simply take restriction fragments of a certain size which were scattered throughout the genome, but they sampled a very small subset of the genome, and we generated SNPs across those regions. A little while later the draft genome sequence had been accumulated all the time and it suddenly became possible to essentially not just compare all the sequences to each other in the defined regions that we had created, but also to look right across the genome. One read was suddenly enough to call a SNP because it could be aligned to this wonderful, free, innovative draft genome sequence and variants could be caught all over the genome with much greater efficiency. Then I think the whole question of developing a dense SNP map became much more a reality.

Yes, I was quite involved in the SNP consortium, probably actually not at the beginning. I think at the very beginning there was clearly a series of discussions, probably founded from the various meetings, to discuss the importance of variation, the recognition, perhaps, that variation was very important not just to genomes and genetics, but to pharma companies as well. So there clearly was an interest in variation within the pharma companies, and I think there had already been some individual investments in individual companies to try to address the problem of getting a reliable, comprehensive set of variants to actually answer questions about genetic predispositions, variability of drug response, whatever it might be. It must have been at least a year, I think, that there was this tremendous negotiation going on between pharma companies, orchestrated, I think, largely by Alan Williamson who is another friend and mentor of mine. Alan must have spent, I think, at least 12 months working on putting together this idea of the SNP consortium, of developing both a public/private partnership and also persuading or encouraging the pharma companies to agree that this particular resource of collecting information, collecting variants, could be considered pre-competitive, and that was a really important element of the program which enabled people to agree that, yes, A, we can't do this alone; B, there's little point in 10 or 11 companies doing the same thing, so let's pool our resources and do it much more systematically in a much more organized fashion, and agree that it is a pre-competitive space. And from that moment on, I think the concept of the pre-competitive space of course then drew on the concept of rapid data release; being able to share the information; different laboratories working to a

common standard, a common set of standards for the quality of the cull; and methods for verifying. Indeed, there was something of a round robin -- not quite a round robin; there was something of a verification process anyway for the SNP culling to ensure the resource was both standardized and high quality. This was the point at which moving from the early negotiations to establish the governance and the guidelines of the consortium to actually getting the job done, and that was of course when the academic labs were approached. Sanger and Washington University and the White Institute, as it was at the time, were certainly approached as -- if we were provided funding from such a consortium, how would we approach to do it? And I think largely independently, two of us at least, if not all of us, came up with similar ideas or the same ideas for an approach, and that lent some conviction to the idea we could all work within our own laboratory organizations but contribute variants of a common standard that were distributed across the genome as a whole.

Well, SNPs can be used in many different ways, of course. Essentially using SNPs for linkage analyses in pedigrees, they have a lot of power over a lot of distance because there are relatively [unintelligible] few crossovers that are informative in terms of identifying or finding your way around the genome and inherited phenotypes with a particular pedigree. But a much higher density of SNPs, essentially the target of the SNP consortium, of course, was stated as 300,000. I'm not sure how much that was genetic theory and how much that was budget-limited, but 300,000 was a stated goal. It was a good density, an average density of SNPs to go for, and clearly did match the ability to start measuring linkage disequilibrium in populations, and this was the start of something much bigger. I think it was always going to be true the more SNPs you have the better you could characterize genomes with using the linkage disequilibrium in different populations, and from my perception, the fields evolved somewhat in parallel. People were working with small regions or, in our case, chromosome. We were looking at LD on chromosome 22 with a set of SNPs and we clearly recognized that calibrated the density of the map at one level, but others, like Jeffrey's [spelled phonetically] in particular, was working at very high density in a very small region of the genome and demonstrating crossovers and demonstrating the ability to really characterize in very fine detail the pattern of recombination in populations. So, clearly, the more denser the maps, the more valuable the resource would be.

I think, from my perspective, I was largely involved in data generation for the actual SNP collections. Two things happened, one quite early on in the TSC, which is when -- I've already described to some extent the fact that actually we could make much greater use of the data we had by incorporating the genome and, essentially almost overnight, we could quadruple the number of SNPs we had in the same resource. Clearly by now the whole concept of collecting SNPs, mapping SNPs, using the draft genome that was getting better and better all the time, of course really changed the game in terms of what we could do to generate a good resource. At the beginning of the HapMap project, or slightly before the beginning of the HapMap project, one of the questions surrounding it -- I remember David Artschuler and Tom Hudson [spelled phonetically] and myself, Mark Daily [spelled phonetically] one or two others, met early one year and discussed the possibilities. One of the elements was actually generating more SNPs, a higher density, and at this point we were able to, once again, actually take a chromosome by chromosome approach and flow-sort [spelled phonetically] more chromosomes and sequence them and align to the draft genome. This was again getting easier, the currency of sequencing was yielding more and more, more efficiently. One of the elements of the HapMap was the recognition that with all the pilot studies and measurements of LD having been carried out, we began to think -- to get a better idea of the

number of SNPs that were required, or the benefits of more SNPs. So SNP generation was actually a first part of the HapMap project as well and was pretty much being done at the time we published the principles of the HapMap project kind of before it started in earnest, but at the same time we were already generating more SNPs in preparation for the HapMap project.

Well, I think much of the leadership of our consortium again came from Francis. I think Francis took a very prominent hand in organizing, and particularly ensuring the involvement of so many people, both who had been involved before and those who were newcomers to it, both large and small. This was the first consortium that worked closely, I think, with Yusuke Nakamura, and Japan was really very much part of this, the single biggest contributor to the first phase of HapMap. That was a very exciting and stimulating entry into the community. And I also remember a very large meeting here in Washington -- I think the Renaissance Hotel or something -- where actually there were members of every possible community and participant and skill set all gathered in this room to really explore over a day and a half, I think it was, what this project was, what it meant. Many different population representatives were present and lots of discussion was had, both outlining what the project would be and how it affected people, how it affected populations, how it affected society, and the impact it could have ultimately on population genetics and disease.

I think pretty early on -- I don't know what stimulated it actually, but I do remember early on I think a number of us felt very keen on sampling different ethnic groups, certainly to pick from Asia, certainly to pick in Africa. I also remember a discussion that then went from that to deciding, and Eric Lander was on the call at this point as well, but a call where we agreed, Eric and myself and several others, I think, that we really needed to go to the indigenous populations to actually collect the samples. While it was much more difficult, it added a whole level of difficulty to the program compared to utilizing an African-American group, a European group, which was already a research cohort or something. But to go to indigenous populations for me felt very important from the point of view of going back to the source. The Africans are in Africa. But, of course, very much also important, it globalizes the whole project. It engaged communities in a way that was at once very challenging, because here were communities who'd never heard of HapMap, didn't necessarily know what the benefit was, and so the idea of contributing freely to a consortium -- they needed to work a lot of that out as part of the engagement. It was a fascinating process. I wasn't very involved in engagement; I was involved in one engagement process in particular with a tribe of the Masai in Kenya, and that was a tremendous experience, and illustrated to me all the work that must have gone into the recruitment that I think was very systematically done by more than one team as part of the HapMap project.

I was enormously impressed by Charles Rotimi's contributions and thinking about the whole thing. There was a simple and very right philosophy, and of course he really, of all people, could do this. He understood exactly the cultural differences and what could and could not be done. I'm delighted at Charles's leadership, which also was happening at the same time as the African Society of Human Genetics was set up, and Charles was the initial chair or president of the African Society. He invited me out there, in fact, which is how I came to be in Kenya, involved in waiting to engage the Masai. Charles was there and they had a small team of us essentially waiting for the call from the local doctor, [foreign language], who eventually drove us out to actually meet the chief of the tribe to discuss what it was we wanted.

And just as you described, of course, we explained or tried to explain the concepts of what we were doing [unintelligible] inheritance, and of course two things emerge from this discussion from a very wise, young tribal chief, a remarkable man, completely on the ball. But we had some trouble explaining inheritance and we talked about things being passed on from generation to generation. This went off in the direction of cattle and goods and things that get passed on. "No, that's not quite what we meant." We then got onto blood, and somehow they understood the concept of things in the blood that get passed on, and they weren't talking infectious disease. They were talking about patterns between generations. They understood then that we wanted to study this idea of things that get passed on, patterns that get passed on, and they were very interested and excited. They got this because it meant a lot of them, the family ancestry was a very strong element for them. So suddenly we got off onto a common ground, common interest, and they absolutely saw why we should be interested in this, and they were delighted to help. The other interesting thing that happened was actually -- whether it was Charles or Pat, I don't know -- asked if they'd like something in return for their generous gift of permission to engage their particular community.

One of the very interesting things, that they -- we asked them I think either Charles or Pat asked them did they want anything in return for their generous gift of giving us permission to engage the tribe. And they -- he came out with it straightaway that the chief HIV testing. And we asked him to expand a bit on that, and he said -- because we don't have HIV. He said yet. He said but we know it's coming, and when it comes we want to be ready. As ready as we can be, we want to do what we can. And so he was completely in touch with what was happening in other parts of Africa even though it was a relatively isolated tribe. It was a fascinating exchange. And I asked him afterwards as we walked across for a celebratory soda in the local soda bar. I asked him if he felt -- clearly he felt part of his tribe, because he was in charge of his tribe. Did he feel part of all the Maasai, the other tribes? And he said, "Yes." And I said, "Did he feel part of all of Africa?" And he said, "Yes, all of Africa." And I said, "Do you feel part of the whole world?" He said, "No." And that astonished me, because he clearly got way beyond his own boundaries of what he saw on a daily, weekly, monthly basis, and his responsibilities, to Africa as a whole -- as a population. But he said he did not feel the same connection with the rest of the world. And that of course was exactly what HapMap project was actually seeking to define. And at that point I recognized we were really coming from rather different backgrounds, myself, and the Maasai chief.

Yeah, I was not involved in the human diversity project at all, and I didn't really know much about it, but I was aware that as the early discussions about the HapMap project took off, that there was this sensitivity. There was this past history that it was very important to work differently to fully engage communities, and I think that was very much a part of the motivation and the agenda for this very big meeting in this hotel here in Washington, DC. The Renaissance I think. Where we did have members of many of these communities present, to have a voice, to speak out, to voice their concerns and their interests. And clearly that was a very major part of community engagement, and of course community engagement became a big part of the HapMap, and rightly so. And I think I'm very glad when it was done on so many levels. It was done on the levels of going to get samples with the indigenous populations, but using a process of engagements to get them to understand and just explore their interest in that. So I think the extent to which that may have been presaged by experience from human diversity project I think is a very good way of managing the situation as far as I'm aware. The process was extremely successful. I don't know

the details, it was another group that was doing it, but at the same time it was certainly good and certainly I enjoyed the brief passage of committee engagement that I was involved in with the Maasai.

So I wasn't very directly involved in the real definition of the boundaries of what's HapMap-able and what is not. But it was clear the concept was there were regions that were difficult for one reason or another. Either we weren't finding snips in those regions or there was something intrinsic about the structure of those regions or there simply wasn't any L.D., and there was perhaps more recombination in this put two or three of those components together and you suddenly get a region that really defies characterization. Not to say it isn't map-able, but to say that it would involve a disproportionate investment to get across the region perhaps, and we didn't necessarily have good approaches at the time to target every region of the HapMap-- of the genome to ensure that it could be mapped to the same extent. And so as a result we had to come up with a definition of so when is the project declared done? Probably not unlike the human genome in a sense, it was just less easy to define. The human genome is clear. Well it's done -- well it's done when we've done the euchromatic regions, because we know we can't do the heterochromatic regions. So it's done in one sense, it's not done in the sense a cytogeneticist would say where's the end of that chromosome. Particularly the short arms of the acrocentric chromosomes. There are ribosomal genes up there, but we can't say it's part of a 2-megabase contig or something. So clearly the unsequenceable regions of the human genome can be defined based on a lot of other evidence about structures being different, presumably perhaps base composition may different, repetitive sequences. We can see the elements that they're in, the unsequencable regions, and of course we can then continue to begin to work on them. Although they take a lot more investment than the rest of the genome. The value of course perhaps considered particularly for supporting genetics and medicine and much of biology of course, you're in where all the genes are. And hence the euchromatic region emerged as being the primary finishing post for the human genome sequence. The same is true of HapMap. But we need a different finishing post, but the problem was the same. The regions that defied characterization for somewhat unknown properties, and this case of course in some cases the properties were less well known because there wasn't so much evidence to say why was that region different, why was it not actually being populated with markers or why was there no LD across the regions. Perhaps we didn't have the right samples to genotype, perhaps more work could've yielded smaller gaps. But I think therefore it's a rather operational endpoint to the project that got defined by a number of things, including the number of snips and including the number of samples in the different groups, and changing any one of those parameters would make a difference to the unmapped regions. At the same time I think we reckon that the first approximation is to -- when we sat round a table I think at Kolfee [spelled phonetically] harbor actually, and discussed what the bounds of the project would be, we felt that actually to have potentially 85 percent of the genome of 85 percent of the sequenced genome, 85 percent of the euchromatic part of the genome essentially captured in HapMap blocks was a tremendous outcome. It was a good endpoint to aim for, and that we should then direct our strategy to try to cover that, rather than saying we should direct all our effort to the remaining 10 or 15 percent. It was important to gather the majority essentially. The job would not be finished by one measure. This is going right back to the concept of the east genome and Maynard but -- even a product with some gaps is a jolly useful product. Look at the 85 percent you have and not the 15 percent that you don't have. Clearly when you move to applying the results of such a structure to genetic associations to genome wide association studies or whatever it might be. The fact that you only have 85 percent of the genome means that

assuming associations or susceptibility factors are distributed evenly throughout the genome, you will find 85 percent of them based on the SNPs that you have if you have sufficient power in your studies. So it was a pretty good starting point. Look at what you can achieve as opposed to worrying about the 15 percent that you can't get at the moment.

Well I think there were two ways to go from that point. I think one was to continue with developing the concept of the HapMap and addressing some of the original areas that we'd had to put to rest for now, to park, which included more snips, more populations as well. More populations question had been there right from the beginning. Clearly we don't know about the level of similarity or real applicability of one sect to another until we're actually doing other populations, so I think that was a tremendous opportunity to diversify the populations involved in the study. And we're still only scratching the surface as far as the human population is concerned. But the other way to go which became more of a focus for me was the unanswered question of how a good genome wide association studies really were, could be, could they be better. Because there was -- there had for some time been many antagonists. People who didn't believe that the GWAS program was really going to either identify the right factors or the sites identified in the genome wouldn't necessarily provide insights on disease. And so I think once this comprehensive HapMap map is available with the snips to enable it to be applied to case control studies. There was a big unanswered question. Well now we have the opportunity to evaluate once and for all the utility of this approach. And we're either going to prove it or it's going to stop garnering funds for further work. This is a discussion in the U.K. So I think John Bell was involved, Peter Donnelly was involved, and the question was, were we really going to mount a program to evaluate for ourselves what the HapMap would bring to powering GWAS studies rather better? And that really was an interesting transition to becoming a national consortium. This was a welcome trust case control consortium started to utilize the HapMap to take advantage of existing case control cohorts in the U.K. to really start to look for can we now find associations now that we have a HapMap of at least a significant density.

Yeah, I wasn't really much involved in the GWAS studies early on, I was fairly compelled by the fact [unintelligible] association study with deep vein thrombosis and the numbers -- the logic seemed to make sense, but clearly as you scaled up, yes there were shortcomings of some of the studies. And the HapMap of course provided at least reduced the limitations of some of those studies by going truly genome wide, or at least asking the questions of 85 percent of the genome in a fairly well measured way. So that was a positive contribution that the HapMap could make to those studies. And I think it did. I think it stimulated a great many more findings -- findings that were statistically strong, developed new ways of actually analyzing the data as a result, and clearly a great many targets of the genome have been found in various cohorts. It's still a struggle I think for two reasons. One is to actually find the variant that is directly associated with the condition, the causal variant is still a leap from the association. And the other of course is these are common conditions and the multifactorial nature of the program means that even if you found a really strong hit, it's only part of the architecture of particular disease or the cause of the phenotype. And there again is another whole complexity to the study of common disease through this kind of approach. It's a big challenge.

By now I was at Solexa, and very much the sequencing technology development and Illumina has acquired Solexa as well. So not since I was less involved in the actual inception of the project, but

certainly some of the unanswered questions of the HapMap and genetic maps and incompleteness of the coverage which had been posed by these projects were clearly once again being challenged. There was the question now we could really overcome in popular -- at a population scale. The gaps between SNPs. And to start in particular I think the idea being able to capture every variant in a person's genome. Completely change the dynamic between indirect and direct association studies. In the direct association study, you're working with the actual variant that's causing the particular condition. The association signal is going to be much stronger than any indirect association that relies on LD, between the cause and the mutation and the actual marker in question. So the idea of having every variant in an individual was actually a concept that I was thinking much more about, even in the 1000 genomes project. The 1000 genomes project sorts the sequence individuals, but to accumulate all the variance they could find to make a better resource of variation. And that already went a lot of the way towards a much -- to fill in the gaps to really study variation across every part of the genome at least that was covered by sequence. And you could really close the gap between the 85 percent HapMap-ability, and the fact that the reference sequence was covering 98, 99 percent of the euchromatic sequence. So what is in that difference between 85 and 99? Does the 1000 genomes project provide those variants? I think that was a very important question to -- for the 1000 genomes project to answer.