My name is Charles Rotimi. And I'm the Director for the Center for Genomics and Global Health. I'm also the branch chief for the cardiovascular metabolic and inflammatory disease branch, which in the NHGRI intramural program. So I direct a team of multidisciplinary investigators who are interested in understanding, you know, disease distribution in different populations.

I was born in Benin City in Nigeria, where I did most of my growing up. And I went to my high school and initial training as a biochemist at the University of Benin in Nigeria before migrating to the United States for further studies.

After graduating in biochemistry from the University of Benin, then you are required to do one year of national service, which is basically a combination of military training and also serving the nation. So after I completed that one year, I applied to universities in the U.K. and also in the U.S. And interestingly, I was accepted into the University of Manchester in the U.K. to study petrochemical engineering. And I was also accepted to the University of Mississippi, Ole Miss, you know, in Oxford, Mississippi to again continue my biochemistry in terms of food nutrition.

So it was a very interesting thing. I presented both to my parents. I'm from a very humble home. So I wasn't at all sure that my parents would have the resources to send me abroad for further studies. So I was pleasantly surprised when my mother, who's a trader, said she could afford to pay for one year. It turns out that the U.S. university fees was slightly lower than that of the U.K. So that's actually what made the decision for me to come to the United States instead of going to the U.K. for petrochemical engineering.

So one year school fees was paid for me, and I was basically on my own after that in the U.S. So my initial socialization to the American culture was actually at the University of Mississippi, where I got my initial Master's degree in healthcare administration before going to Birmingham, Alabama to study Epidemiology.

My doctoral thesis really was actually quite interesting. It's quite removed from what I'm doing now. It was actually looking at the impact of working in the foundry and engine plant at Ford Motor Company foundry and engine plant in Cleveland, Ohio, in relation to lung cancer and stomach cancer. So the union then was a little bit worried that the hourly workers were experiencing a higher rate of stomach and lung cancer. So my supervisor there at the University of Alabama were given the contract, you know, to study, to determine if this is really true and if this excess is actually due to occupational exposure or something, you know, that people are doing at home, or some combination of these. So that was what my PhD thesis was on.

I was basically using epidemiological methods to look at records and determine what kind of exposure -- first of all, determine was there increased risk, and if there was, why. And it was a very fascinating, you know, study.

Again, my story is one of those that really highlight the fact that there is really no one path, you know, to what you're going to become in the future. Right after graduating from the University of Alabama, I got a post to a position at Loma Linda University in California. So then I had a little Honda Civic that I drove across the country because I couldn't afford, you know, the other modes of transportation, to start my post-doctoral work again with my family. And that was actually very

fascinating. I was working with a P.I. there who was studying Alzheimer's disease, and that's what we were working on.

One of the things I love about training epidemiology is once you understand the fundamental methods, you can actually apply to any disease. You know. So it was quite comfortable for me to be able to do my post-doctoral work within that Loma Linda University, which is a very, very wonderful University. But I think probably the more interesting part of that study is really while I was there. I saw this ad one evening in the paper by Richard Cooper, you know, that says that they're looking for somebody who would be interested in studying hypertension, you know, in the African diaspora in different African populations. And that he was interested in an epidemiologist, you know. So the way the ad was written, it was absolutely very fascinating. So I called Richard, and I said, "Richard, you must have written this ad with me in mind," I said, "because this is exactly what I want to do." You know.

So he, you know, he invited me for an interview. And I accepted the job. So I didn't complete my post-doctoral work. So I moved from Loma Linda to Loyola Medical Center in the suburb of Chicago. And that's when I started really, you know, studying and understanding health disparity, and how to use the migration and design of the African diaspora experience to shed light on why diseases vary as you go from rural Nigeria, or rural West African countries, you know, to the Caribbean and to the U.S.

When we started, we really wanted to understand, first of all, what is the prevalence of hypertension as you go across these different populations who share very recent ancestry. You know. So we wanted to understand that. And then once we understand what the prevalence was, wanted to see if we could shed light on some other factors that are driving this, you know, different prevalence as you go across this different African, you know, populations.

So we put together--again, Richard wrote the initial grant, you know, to do this, and it was to enroll over 10,000 individuals. And we were quite happy to have been able to do that in a very, very rigorous way. We standardized the procedure, because blood pressure is one of those things that vary quite a bit. So we needed to make sure that we're using the same procedure across the different sites. Otherwise you wouldn't be able to compare results. You know. So there was a lot of standardization that went through that process. And we were able to come up with data for over 10,000 individuals from Nigeria. We have opened rural sites in Nigeria. And in Cameroon, we also had opened rural sites. Then in the Caribbean we have St. Lucia, Barbados, and Jamaica. And of course African-Americans in the Chicago area of Maywood, Illinois.

So the finding, you know, from this study really shed a fundamental insight into the impact of the environment that people find themselves on blood pressure and hypertension, you know, distribution. So it was very clear that it varied from about seven percent in rural West Africa to about 16 percent in the urban centers, like in Bardo and Lagos, and then to about 26 percent among the black nations of the Caribbean. And over 34 percent among African-Americans in the Chicago area.

So you had this very almost monotonous increase in blood pressure and hypertension prevalence as you go from rural Africa, you know, to urban, you know, Chicago via the Caribbean. You know. So it was very, very fascinating, you know, for us to find that.

And then we shed light also on that that gradient that we were able to establish was due -- we were able to explain close to 60 percent of that variance by looking at things like salt intake; you know, how heavy you are; you know, and also level of physical activity. You know. So you see the impact of the environment was really, really well demonstrated, you know, in that study. Yeah.

That does not mean that there are no genetic susceptibility. But if you are actually comparing and trying to understand a disparity, then I think you would get a much bigger bang for your money by looking at the environmental factors that drive these conditions, yes.

I think that's a very important question, because one of the major challenges in doing even genetic studies is the ability to characterize the environment. I would say that at some point in all of this, even in genomic studies, the real limiting step is really going to be the ability to characterize the environment. Because that is what is always changing, you know, and you have to be able to do that very well. So for us, one of the things that we were able to do was to measure things like height and weight. You would think that measuring height would be a very straightforward phenomena. It is absolutely not. For example, during one of our studies, we lost a lot of data points just because the interviewer who was measuring individuals did not tell people to straighten up against the wall so we can use this stadiometer to measure their height. So some people slouch. So a lot of people were very short, you know, for their weight. So their body mass index was really high.

And then you have some women -- because in Nigeria women wear headgear -- and the headgear can be sometimes up to six inches. You know. So some of the interviewers did not tell the women to take off your headgear before they were measured. So they were much taller than what their true height is. So all of those cause major problems, you know, in terms of your ability to measure. So things as simple as height can get very complicated if you don't standardize the procedure and you don't train very well. You know.
And there are some environmental factors that are really difficult; things like diet. For example, when we wanted to characterize salt. We felt that our best measure, because we are not dietitians, our best measure would be to look at it from a biochemical parameter point of view. So we collected urine samples. And from the urine samples we were able to measure, you know, in the sodium potassium. And that gave us idea as to consumption. So we were using this secretion to approximate intake. You know. So you have to be clever about how you do some of this stuff.

And there were collateral things like education, income, occupation, which again are very, very, you know, important.

In terms of the consent aspect, one of the things that we did very well is that we engaged the community and also the community leaders. And we made it clear what our intentions were and why we were doing what we were doing. And, you know, to let people know that for the very first time we are contributing to people's understanding of why some people get high blood pressure and others don't. And how do we study this at a population level so that we can make

recommendations in terms of preventive strategy. And also for people to just be aware of the fact that the fact that you don't feel anything doesn't mean your blood pressure is not high. You know. So by doing that, I think we brought the community along.

And another plus in doing studies in places like Nigeria, Cameroon, you know, and the Caribbean is that people are still very receptive to research. So it's not too complicated to get people excited about, you know, doing this kind of study. So that helps in our ability to be able to recruit people. And then another major factor is we employed people who actually live in those communities. So the participants in this study know the people who are working in the clinics and the people who are knocking on their doors and asking them to come participate in the study. So that was also very, very useful, to have familiar faces out there, you know, who people can relate to. Yeah.

Yes, there have been some very bad history in terms of biomedical research in the African-American experience. And therefore, people are skeptical, which is I think rightfully so, that people should be skeptical. But one of the things that we do -- I do in my study is basically to train my [unintelligible], you know, research, you know, assistants who go out in the community to be aware of this history; not to run away from it; understand it and explain it to people. And more importantly, let people know what are the things that we have put in place to make sure that the likelihood of something like Tuskegee occurring is much more reduced now. You know.
So I think for me it's important to acknowledge it, but to also let people know that that should not be a reason not to participate in biomedical research, because the community will be the loser in the long run. But what you need to do is when you are participating in research, have your eyes wide open. The fact that I have a dark skin like you doesn't mean that you should trust me. You should still ask questions just like you ask anybody. And make sure that the right things have been put in place. First of all, you make sure that my study has been approved by the appropriate authority. And that you have somebody to call if you feel that anything is going wrong with the study that you don't like. And that person should be independent of the study, itself, or the study personnel.

So we explain all of those things, and we let people know why they need to participate in a study like this, especially in genetic studies.
You know, when you feel uncomfortable at any point in the process, stop. You know? But if you feel that the study is going on in the way that you like, you should participate, because that is the only way we can understand what is going on in your community. Yeah.

One of the things that I also observed, you know, with my mentor, Richard Cooper, when we were doing all the studies is that when you see somebody who is diabetic or hypertensive, they tend to have the other condition. So if it's somebody who is diabetic, they tend to also be hypertensive; they tend to also be obese. You know? So for me, it made absolute sense to try to study these conditions. So I started looking at my whole research, not in the context of these very specific diseases, but looking at it for metabolic disorders. You know? So bringing them together, you know, so I looked at hypertension, obesity, and diabetes as a triangle. You know? And they feed into each other. And our ability to understand how these things are related I think will shed tremendous light on how we prevent it, how we treat it, and how we communicate to people about

their condition. You know. So to me that was absolutely the way to go in terms of my own research, you know, in the research activities.

So I started with Richard Cooper with hypertension. We saw the impact of obesity. Then during that process, we started talking from initial discussion about diabetes, started actually with Francis Collins and Georgia Dunston . So Georgia Dunston, who is a professor at Howard University, you know, was doing a sabbatical in Francis Collins's lab at the NIH. But before then, Francis had done missionary work in Nigeria, where he treated quite a few diabetic patients. So I think that stuck with him for a long time.

So when the opportunity came about in terms of discussing how can we study diabetes in West Africa, with Georgia and his lab, so there was a discussion about how do we move this forward. So I was brought into that picture as an investigator who is doing work in Africa to see how we can put together a study that would shed light on first of all, what is the risk of diabetes, and what are the factors that are contributing to diabetes from the environmental point of view, and also from genetic point of view.
So that's how the diabetes component, you know, started in a sense.

At that time I was already very, very interested on the genetic contributions, you know, to these various diseases, because we studied from the EPI point of view, and we saw that there was still some things that we noticed. For example, why is it that you tend to see hypertension run in certain families; and diabetes right in some families and not in others. So there was an aspect of what we were studying that was not completely explained by environmental factors. So I wanted again to bring genetics to bear on that. I think to cut a long story short, I accepted, you know, to come to Howard University to lead the genetic epidemiology unit. And I got to be the director of the genetic -- so the genome center there, you know, had four major areas. It had genetic epidemiology. It had statistical genetics. That was led by George Bonney. And it had molecular genetics, which again was led by Georgia Dunston and Rick Kittles. And it had ethics that was led by Charmaine Royal . And it was a very, very successful center, you know, then. It attracted a lot of research funding's. And we did quite a bit of what I'll consider very good work in shedding lights on genetic and environmental determinants of different diseases: Cancer, diabetes, hypertension, you know, across the African diaspora.

Again, there are several toolkits. And I think depending on your approach and how you want to -- and also the state of technology tends to be very, very critical, because when I started out, you know, in terms of trying to look at genetic, you know, contributions to these diseases, at some point we were able to look at one gene at a time, and, you know, do some, you know, basic, you know, snips, you know, in these genes. And, you know, we call that a candidate gene approach, you know, then.

And then of course with more markers, you know, come in as a result of things like the HapMap, and we were able to do linkage studies. You know, which, you know, basically, you know, helps you to determine regency in a chromosome that is tracking a disease, you know, in families, you know, basically. And again it's based on the principle that things like that are related together, that are close together on the chromosome, and that stick together during miosis, you know, in the

sense they are not broken up. So you can use the marker to track, you know, that chromosomal region. And if you see a signal, then you could do more work within that region to see if you can't try to localize the specific gene.

So there's the linkage study. And it is the association study also, which basically, you know, you're looking at people who have the disease and people who don't have the disease. And you are trying to see, do I see an increased risk when I look at these people. And all decreased risk can be susceptibility or resistance. And then you –- basically that is based on having a large number of people who are cases, and a large number of people who are controls, and you do your association study and you use statistics to see what is the relationship. And see if that relationship is significant. And then you make your inference.
But all of these, you know, association studies became actually practical as a result of the sequencing of the human genome and the subsequent spinoffs like the HapMap and the Thousand Genomes. Because we couldn't do it, you know, cost-effectively enough before all these things, you know, came about.

So these are some of the tools that we have, you know, to map, you know, these various diseases. The complex diseases, what we call complex diseases, or diseases that are non-mendelian which basically as a result of much poor genes and much poor environmental factors, are pretty difficult to track, you know, because you're not just -- it's not just one gene. And it's not just one environment. It's a constellation of different things coming together to increase risk or decrease risk. You know. So it's quite challenging, you know, process. Yeah.

The challenge, you know, is really how do you begin to tease all of this apart in a way that you actually are able to quantify your risk, my risk if we carry these markers. And maybe in the end design drug targets, you know, for these various, you know, genetic markers in a way that you can now try to treat people, or come out with better medications.

It's a very, very difficult thing. You know, we all know, even at our own personal level, we all know sometimes you know the reality, but you just are not able to implement some of the changes to take advantage of the information you're getting.

But I think the basic thing is letting people understand the concept of risk and the concept of probability. Okay? And that it is a chance. You can come from the various activity that people engage in on a daily basis. You know, that if you have a deck of cards, and you have more of the cards that are similar. The likelihood of picking that card that are more than one in a deck is higher. Okay? So if you start with that concept, so basically what you're saying is when we find people who have 10 types of this type of genes, their risk of diabetes is much higher. Okay? And what that means, it doesn't mean that you're going to get diabetes tomorrow. What that means is if they put in place, you know, things like reduce your weight, do more physical activity, watch what you eat, that they might indeed be able to overcome that genetic load that they carry. But it's something that you have to do for the rest of your life. It's not just do it for one month and you can give it up. But you have to maintain an ideal weight and, you know, eat properly, and do your physical activity.

But we also have to be realistic when we're communicating that, and tell some people that if you have a very strong family history, that you may be able to postpone when the disease will occur. But you may not be able to prevent it completely. So we have to go through all of these dynamics and all these ways of explaining risk and probability to people in a way that people can understand and implement it in their daily life.

So I think it is an issue really of education and communication. So it's not just the scientists at this point. They are to bring in people who actually have the skills to be able to communicate, you know, the risk, you know, to individuals. So it's, again, a very daunting process. Yeah.

The genome wide association study, what we call GWAS, is really I think one of those fundamental tools that have really changed biomedical research. You know, in a way of looking for disease-gene association. It's really revolutionized, you know, what we can do. And that came about because of our efforts in the HapMap Project, the International Habitat Mapping Project. Because before then, you couldn't really interrogate the gene. It was too expensive, you know, to do. It wasn't practical. And we really didn't even have an idea of all the markers that it could use to do that kind of study. You know.

So investing in the HapMap Project at the national and international level really led to the genome wide association revolution, in a sense. And it also led to the ability of the biotech companies to be able to develop the genotyping arrays in a way that it could be used from one study to another in a more cost friendly, you know, manner.

I was involved in discussions right from the beginning. And we recognize that, again given the revolutionary history of humans, that, you know, we all started -- our ancestors started somewhere in Africa and before we migrated to different parts of the world. And because of that, you just have more variation, you know, in the African people. So it made absolute sense that if you were going to characterize human genetic variation, that you have -- that the African population is represented in that effort.

So right from the very beginning there was a discussion to at least represent some of the continental diversity that we currently have, given the limited resources. We all knew that that wasn't the complete answer to all of this, but we needed to start somewhere. And those four populations, you know, the Yoruba and Japanese and Chinese and also European, were initially put together, you know, in terms of the HapMap Project. So I was involved in that discussion of which population to sample from Africa. So I was the P.I. of engaging the African community for the Haplotype Mapping Project. And I did that with my colleagues. We were able to engage the Yoruba community in the Benin area in Nigeria to participate in this effort. That was my very first full committed engagement activity, to really tell people why they need to participate in this study and to hear from them why they may want to or why they may not want to participate in that effort.

So again, right from the very beginning I think we were very, you know, cognizant of this various aspects of the project that needed to be brought to bear.

And one of the distinctions between the HapMap Project and the diversity project is really the fact that we wanted the HapMap to be biomedical research based. It was health. We were designing a tool to help us understand human health and disease. Not necessarily what I'll call the other gains came later in terms of population genetics, all that. But the main focus was how do we characterize the human genome in a way that it will become a tool for us to be able to understand health and disease. That was really the driving force behind. And that set HapMap aside from all of the diversity projects of the region project that was going on at that point in time. You know. And also that also helped us when we were doing the community engagement to be absolutely clear, you know, to people why we're asking them to participate in this study, and where their resources would be useful.

The process was really identifying, you know, key investigators who, you know, work with me on this engaging the African community. In the Benin area, for example, engaging the Yoruba, the local P.I. for that project then was Clement Adebamowo who was the professor of surgery at the University very close to the community where we were doing this work. So that was a key person. And then I had Charmaine Royal and also Patricia Marshall who were ethicists and knew, you know, social scientists who knew how to engage, you know, communities. You know, so it was bringing these forces together that we approached the community. First of all, we identified the ethos of that community. We identified the relationship structure of the community. And then we approached them with the help of Clement Adebamowo. And, you know, we told them about what the study was all about and what we want to put in place to make sure that, you know, individuals who participate in this study are protected, and for them to actually understand what they're doing. And the informed consent that they are going to sign, how the resources are going to be used, how they are going to be broadly shared, and also the fact that no identifying information would go with the samples. You know, even if the consent forms that are signed would not come to the U.S.; it would stay in Nigeria. And so those samples could never be linked, you know, to an individual in a sense.

So we went through all of that community engagement, letting people know what this study is about. To our own fascination, when we started talking to people, we were quite surprised as to the level -- scientists tend to want to underestimate the level of understanding of community members, and I think that's a big mistake. You know, the fact that somebody cannot say haplotype does it mean that they don't understand what you're doing. You know. So we were very, very surprised as to how people were able to relate back to us what did they understand the HapMap was going to do.

I remember very clearly one of the comments made by one of the ladies who had been participating in this community engagement was the fact that she said this will enable us, you know, understand how people who left Africa are related to us now. You know. So that was a fundamental way of really articulating what we were going to learn from the HapMap. To me, I was completely blown away. When somebody we thought, you know, didn't understand everything can actually put in such simple language. So we started using that expression ourselves, you know, to explain the HapMap, you know, to people. So it was absolutely quite fascinating. The whole community

engagement was quite interesting. And the only other thing I'll say is don't engage a community if you don't want to hear what they have to say, you know, because communities can tell you things you don't want to know; that you didn't think about while you were designing your study, and some of the concerns that they may have, you know, about the study.

So when you are doing community engagement, you have to have an open mind and be prepared, you know, to deal with issues that you did not think about as a scientist.

She was really the overall coordinator from the NIH point of view. And, you know, Jean McEwen and also Lisa. You know. So Jean especially went with us, you know, much more times to Nigeria. She participated in the community engagement aspects. So she brought her expertise from the ELSI, you know, the ethics point of view. And she had a wonderful understanding of the informed consent process and what are the things that need to be put into that document to make sure that people really understand and appreciate how their resources were going to be used.

So Jean was there again when we were engaging the community right from the beginning to the end. And she was absolutely invaluable to the whole process. You know, in terms of how do you do this. You know. How do we put this in place in a way that we can actually call this community engagement? And I think HapMap set an example that I think is really difficult to beat, because up to now the CAG, community advisory group, still received documents from Coriell biorepository repository in New Jersey about how those resources are used. That's unheard of in biomedical research. So the standard that HapMap set is extremely high, and they have lived up to it so far.

I regard Francis as a mentor, you know, basically for my own career development. And I have always, you know, seek advice from him and talk to him about things. And we have been engaged, again, with the diabetes study and also of course with the HapMap and the Thousand Genomes. And when I was coming to the NIH, it was basically discussion between myself and Francis in terms of setting up the center at NIH. And then we brought in Eric Green then, who was scientific director at that point in time.

So I really do see Francis as a mentor in my own career development. And of course he's one of the brains behind, you know, the sequencing of the human genome and also the HapMap Project. I think he had a vision about the fact that we can actually characterize imaginative variation and use it to understand human health. And that, I think, was the driving force, you know, for Francis.

Right. Again, I think Francis has, you know, pretty broad knowledge base. And he appreciates the fact that you have Mendelian diseases and you have, you know, complex traits. And his own interest in diabetes really again is illuminating in that thing in the sense that these are, you know, diseases that you need genetic epidemiologist, you know, in the sense of bringing in the environment and bringing in the genetics and try to understand how these things influence each other, you know, to increase risk of disease.

So we really talked from that point of view of, you know, complex traits and also from the point of view of making sure that genomics is a global exercise, not just an exercise of rich societies. You know. So Francis is absolutely passionate about that, and that is my own complete passion in the sense that whatever gains we are going to get from genomics, that we make sure it's global; that it applies to all human populations.

And so, for example, when we were starting to do, like, the HapMap Project, you know, discussions with Francis on that, we could have sampled Yoruba people in America. Okay? It would have been much cheaper; probably much faster. But I was absolutely opposed to that. And Francis appreciated that completely; that you cannot do something like the HapMap Project if you don't go to where the people you are sampling actually live. You know? And how can you actually engage their community and say you have done a community engagement when you are doing the engagement in America, but their home is thousands of miles away.

So it was pretty clear during our discussion, you know, Francis appreciated that very well when I indicated that we have to go back to Africa to make sure that we are trying to do the community engagement and let people know what those resources are going to be, and engage Africans on the African continent, not in other parts of the world. Yeah.

Yeah. Again, my take on understanding human variation, period, is we have to sample as many human populations as possible, whether they are small or big. I think we have to cover the breadth and scope, you know, for us to actually fully appreciate what is the complete picture in terms of human genetic variation. So the more population we sample, the more population we genotype or sequence, the more we are going to know about variation.

But we also know because the human history is a very, very recent history in relation to other organisms, you know, or at least in relation to the planet, you know, itself, that we are very, very recent. And because of that recent history, we share a lot of our variations. So very few human populations can adequately represent, you know, the variation up to 99.5 percent of the time. You know. We can capture that. But that's, you know, .5 percent or so that is left tend to harbor quite a bit of some of the things that we are interested in in terms of health. You know. So it makes sense to sample people not from the point of the socio-descriptive, like race or ethnicity, but from the point of view of history; of human history migrations; where people have lived; where are the sources of variation likely to have occurred from: Diet, climate. You know, those are the things that shape the genome. And sometimes those things coincidentally overlap with our social descriptive of ourselves. But that's not the evolutionary purpose. You know?

So for example, if you're interested in kidney disease and the APOL-1 gene, for example, you will only see it in the part of Africa that was endemic to Trypanosomiasis; the so-called African sleeping sickness. You don't see it in a large part of Africa, and you clearly don't see it in Europe or Asia. You know? So it's not even a question of doing Africans. It's a question of doing that part of Africa where they needed to survive that kind of very dangerous disease. You know.
So for me it's very, very important for us to appreciate why we do what we do, and not to overlap it with our social notions that may distort our understanding of human genetic variation.

Yeah. Again, but the whole concept of, say, the rare variants, for example, I think if you push that, and push it, and push it to its limit, it becomes an individual project. You have variants that you may carry that your parent may not carry. You know? So where do you stop and where do you slice this, you know, I think becomes, you know, some of the things that you have to consider.

Clearly the more populations we sequence, the more diverse population we sequence, the more we're going to pick up on rare, and rare, and rare variants. There is absolutely no doubt about that. But we're not doing that from the notion of race or ethnicity. No. We're just doing it -- why do we have rare variants? It started from a place. It has not had time to spread. Okay? And if that population mates with another population, you know, you may carry it on and it may start to spread.

So something is rare to the point of –- it's a time thing. You know, the more time, you know, passes, the more likelihood that that thing will spread, you know, toward the past. If there is no gene flow restriction.

Again, for the phase 3, we wanted to make sure that we expand beyond just the continental thinking for the phase 1, you know, that led us to identify a little over a million, you know, markers, you know, in the sense. So we wanted to expand it to other such populations globally. Okay? And some of it was convenience. You know, some of it was scientifically strategic in a sense. But really it was to try to say how do we capture more populations from Europe. You know. How do we capture more populations from Asia? How do we capture more populations from Africa?

For example, the African part, you know, that I was engaged in, we wanted to make sure we capture individuals that are from different parts of the language group, the major language groups. You know, for example, the Bantu expansions. You know, the Nilo-Saharan. You know. Wanted to capture so the Maasai, you know, the Luhya, the Yoruba. We wanted to expand that in a way that we know that some of the evolutionary history of those various populations may have introduced differences that needed to be captured, you know, in a project like the HapMap. You know.

So it was basically trying to involve more global populations in a way that we can capture more of the variations. Yeah.

I think for me one of the major things that HapMap showed to us is that indeed we are very, very similar as human beings; that we share a lot of things. And that there are some, you know, things that we don't share. And when we quantify that, it turned out to be about 10 million in a single nucleotide polymorphism that tend to differ, you know, between no two individuals. So when you compare individuals. You know. And that those differences tend to be important when you're talking about health and disease. You know, in a sense.

So HapMap help us to understand in a very clear way that a lot of things that we share, and things in the variation in the human genome, is shared. You know, what we call the common variants. They are common, you know, because we share them. You know. And it's also verified that human history is indeed very recent. So we still share a lot of the variation that we see. So HapMap was very instrumental in communicating that message.

It also enabled us to now have a map of where these things are located in the genome. And in a way that we can genotype them in a cost efficient manner. And for the very first time, scientists were able to interrogate the whole genome instead of specific genes in relation to human disease. That really was indeed fundamental in biomedical research.

Absolutely. You know, in either the beginning and, you know, the major thing that made GWAS, you know, happen. You know. It wouldn't have happened without the HapMap or a similar project, you know, to generate that kind of database that the HapMap, you know, generated. And HapMap also gave us understanding of, you know, signals of natural selections, you know, where you could actually look in the genome and you see where the environmental contest had shift the genome in a way that is specific to that part of the world.
You know, again, we get very good knowledge, you know, from the HapMap. For example, about something like lassa fever. Or even, you know, skin color or hair color. You know, those things were indeed adequately characterized by the HapMap Project, yeah.

What it has given us is we were able to sample populations that were not part of the HapMap or the Thousand Genome. Okay? So the very first contribution from the African, you know, genome variation project is really expanding the level of diversity that we now have in terms of samples and data and understanding of variation across the African continent.

So again, one of the fundamental things that we need to appreciate and appreciate fully -- and we're just beginning to appreciate it very well -- is the huge diversity on the African continent. And we needed to sample more and more from the African continent to capture this diversity. Okay? Because of language and barriers, you can travel, you know, 200 kilometers and be in a completely different environment where you don't understand wherein this person is talking in terms of language. And nothing restricts, you know, gene flow, you know, better than language in a sense. So there's a lot of characterization. And of course humans have lived the longest on that continent, so have had opportunity to have more variation in the sense.

So by sequencing and genotyping more African population and putting that in the public domain, we are contributing to the efforts of the HapMap and the Thousand Genome in a way, you know, that will serve understanding of disease; genetic basis of disease on the African continent

So that is what the variation project is really about, was to expand the African populations that are now available, publicly available database. And then we were able to show also signs of selections. That resource will also contribute to genotyping, the GWAS array that we are developing for the H3Africa, or doing research in Africa.

One of the things that has been very troubling to people like me and people who study African populations is that the initial generation of GWAS chips are not very efficient for interrogating African genomes. Again just because most of the variation, and most of the characterization came from European and Asian ancestry population. So they were not very efficient for African populations. So by characterizing more African populations, and then making those data available, and thereby biotech companies are grabbing these variants and they're putting it on the chip, we are now having more efficient African chips. Some of them, newer generations, what I'm using

in my lab, you know, as we speak, you know, to look at diabetes and other diseases. So we are making these tools more efficient by genotyping more African populations. Yeah.

The genotyping for the African variation project was actually done at Sanger, in the U.K. Okay. And with our colleagues there at the Wellcome Trust Sanger Institute in the U.K. But the H3Africa Project at the Humanitarian Health in Africa Project is actually put in place in infrastructure to be able to do this kind of genotyping now in Africa. And you are beginning to see evidence of that. So not just finding disease chain, but also creating the opportunity for people to understand the technology and be able to use it, you know, locally, create, you know, incentive for jobs and also for training of young men and women in this technology.

Yes. Again, the fundamental vision, the principle behind the H3Africa is really to empower African investigators to do research about conditions that are important to the continent, and to do that research locally. And then to be able to analyze it and write the papers. And then increase the ability to compete for subsequent funding's, so that biomedical research can go to it in a much higher level on the continent.

So that is really the driving principle behind the H3Africa. So before H3Africa, what you tend to see was African investigator, they don't talk to each other. They talk to people in Europe, to people in America or in Asia, because they were basically following the money. Now the H3Africa is funding investigators universities in Africa directly. So although they can collaborate with whoever they want, but the money goes to the African institution and the African investigator, and so therefore empowering them to drive the fundamental questions that they want to address; that they think is important to their people. Yeah.

Yes. The Common Fund is really -- it comes out of the director's office, you know, from Francis Collins's office, or whoever becomes the director of the NIH. And this basically direct is designed to help bring novelty fund projects that would otherwise not get funded through that mechanism, but to catalyze, you know, biomedical research in a way that is not easily done through the other, you know, mechanisms.

So it's a very, very wonderful idea. So that's how H3Africa is funded, because it's really putting aside, you know, a sum of money allowing African investigators to come up with their own specific questions and apply for that resource. So the Common Fund has been instrumental, of course along with our colleagues from the Wellcome Trust in terms of their contribution to this effort.

So between the Wellcome Trust and the Common Fund, the H3Africa has indeed really revolutionized genomic studies on the African continent in a way that would have been otherwise very, very difficult to do. And, you know, that's where discussion started up, you know, with me and my leadership with the African Society of Human Genetics, our colleagues there, you know were worried about the fact that African investigators were not participating fully in genomic research. And we wanted to change that.

It is high risk/high reward project. It is, because we really don't know how all of this are going to play out with time. But we are hoping that the initial success stories seems to be pointing in a direction that this is going to be more of reward than risk, in a sense. But it was really a leap of faith to put in this kind of resources and say let's go for it, you know, in terms of the African continent.

And I'm very extremely proud of the fact that the African investigators have stepped up and really made us very proud; made this a success story.

Right. The H3Africa has funded multiple types of research. It's funding diabetes research, which is again in almost in 11 countries now. It's funding microbiome projects. It's also looking at cervical cancer on the continent. And it's looking at cardio metabolic disorders, you know, across multiple countries. And there are, you know, conditions like, you know, Trypanosomiasis, and also pharmacogenomics, you know, in terms of how do we better understand how people respond to, you know, to drugs in the African continent.

So there's a really wide, you know, research areas that H3Africa is funding, and we are experiencing some very, very initial good success stories, you know, coming out of this. Yeah.

And it's very important that as we use genetics/genomics to define drugs and design drugs, it is very important that we have to understand the fact that we need to bring all global populations to bear on this. Because we know that for reasons that we've talked about earlier in terms of natural selections, you know, and just being in one location versus another for a long time, structures our genome, you know, somewhat differently, and that we can respond to drugs differently. And therefore, if we don't understand the global implications of human genetic variation in relation to that drug, then we may not fully appreciate the side effects, or even the efficacy of such drugs as we go from one community, you know, to the other.
The way I sort of phrase it is will tomorrow's medicine work for all? I think that's really sort of the bottom line. And we have to make sure that it does; you know, that we can make sure that tomorrow's medicine will work for all human populations. And that when it comes to pharmacogenomics really, I look at it as going to the tailor. If you want your clothes to fit, you have to present yourself to the tailor to be measured. If you rely on my measurement, guess what? Your clothes is not going to fit you properly. That is really the bottom line with pharmacogenomics and with the whole concept of precision medicine. That we have to characterize the human population as deeply as we can. And at some point, it has to be at the individual level.

I think the Thousand Genomes basically built on the HapMap. The HapMap we use new technology to genotype, you know. Basically we are identifying things that cross the genome. Whereas the Thousand Genomes use a combination of these factors, including sequencing. And basically that allowed us to identify things that we couldn't pick up with the HapMap. And so we have more detail, you know, fine-grained characterization of human genetic variation. Instead of 3.1 million, we are talking about over 80 million variants being identified with the last publication of the sequencing of over 2,500 individuals with the Thousand Genomes.

So again, all of this -- one of the things I've appreciated about the genome efforts are starting from the sequencing of the human genome is really the appreciation of the fact that we need to start somewhere and then build on that and build on that, and continue to build on this. Sometimes some people see that as a criticism. But I think I see that as a realistic way of approaching a major, major biomedical initiative. And I think the success so far seems to have supported our strategy of continuing to build.

So we said Thousand Genomes, but the [unintelligible] actually sequence over 2,500 individuals, because again, we know the more we sequence, the more we are going to be able to capture these variations; especially the rare ones. And they have a much broader and deeper final map of the human genetic variation.

So again, one of the things that the Thousand Genomes is doing for us even in the context of GWAS, is we now have a much more comprehensive reference panel for imputation. What do I mean by imputation? That is basically when we do [unintelligible], and we use this genotyping arrays, you know, there are gaps. Okay? So in a typical one, they give you, like, 2 million, you know, snips. But if you do imputation, and you fill in those gaps using that reference panel, you can get up to 20 million. And that just increases your power to find genes, you know, in relation to disease and health.

The bioinformatics, I think without that we would be nowhere, really, because you can sequence all you want. If you don't have the tools and the know-how to interrogate that sequence and characterize them in a way that people can interact with it, you really all you have is just a mess of data on your computer. So the bioinformatics tools that were developed from the HapMap, the Thousand Genomes, has really, really enabled us to have a much better appreciation of human genetic variation and how to relate that, you know, to disease.

So that is fundamental delivery when it comes to the Thousand Genomes or the HapMap, you know, Project. That bioinformatics strategy and, you know, real strong competition of biology understanding and very clever people, you know, who are able to interrogate and design, you know, software and help us to interrogate this kind of huge data that you really cannot -- it's not your typical Excel spreadsheets, you know, that you put this kind of data on. So you really need people who understand bioinformatics and also, you know, competition in the work, you know, how to design software.

It's really, you cannot over emphasize, you know, the role of a bioinformatician. It's central, you know, to all of this. Otherwise you just have a blob of data, and you can't make head or tail out of it.

It's the annotation of that data; making sense of that data; and putting it in a way that people can use it. Otherwise, you know, if the HapMap data wasn't friendly, nobody would go there. Nobody would use it. So to be able to structure in a way that somebody in Africa or in China can log on and be able to download that data and be able to use it with clear documentation on what has been done is just absolutely fantastic.

You know, for me I think it's important for us to appreciate, you know, projects like HapMap and Thousand Genomes and African Genome Variation, and, you know, H3Africa because of what they do. They tend to revolutionize our ability. They give us tools that we otherwise not have if we don't have those kind of international effort. People coming together from different parts of the world to say we need to develop a human good, you know, that everybody can use to understand biomedical research.

Absolutely. I think it will be tragic if we go through all of this and we exacerbate disparity instead of making disparity go away. You know. So I think by representing different human populations and being careful about the way we interpret the results, I think we will, indeed, shed light on human history, how we related, and human health, you know, in a way that we couldn't do before.