Informed consent for: "The ethos and effects of data-sharing rules: Examining the history of the 'Bermuda principles' and their effects on 21st century science"

University of Adelaide

Duke University

Researchers at the University of Adelaide, Australia, and the IGSP Center for Genome Ethics, Law & Policy, Duke University, are engaged in research on the **Bermuda Principles** for sharing DNA sequence data from high-volume sequencing centers. You have been selected for an interview because we believe that the recollections you may have of your experiences with the International Strategy Meetings for Human Genome Sequencing (1996-1998) will be interesting and helpful for our project.

We expect that interviews will last from 30 minutes to much longer, but you may stop your interview at any time. Your participation is strictly voluntary, and you do not have to answer every question asked.

Your interview is being recorded and we may take written notes during the interview. After your interview, we may prepare a typed transcript of the interview. If we prepare a transcript, you will have an opportunity to review it and to make deletions and corrections.

Unless you indicate otherwise, the ***information*** that you provide in this interview will be "on the record"—that is, it can be attributed to you in the various articles and chapters that we plan to write, and thus could become public through these channels. If, however, at some point in the interview you want to provide us with information that might be useful for us to know, but which you do not want to have attributed to you, you should tell us that you wish to go "off the record" and we will stop the recording. We will, however, take notes for our own use. When you are ready to go back "on the record," we will resume recording. Anything you say while "off the record" will not be on the audio recording and therefore will not appear in the transcript.

All ***materials*** from your interview (audio recording; transcript; interviewer's notes) will be available only to members of the research team affiliated with this project, unless you consent to their wider use, as described in the paragraph below. The digital materials will be maintained in a secure, HIPPA-compliant drive at Duke University. The paper materials will be stored in a locked cabinet.

In addition to the scholarly articles and chapters that we plan to write, we also hope to create a resource for other scholars and members of the public. We plan to post some of our research data to online digital archives. While we will use your "on the record" comments to inform and write our articles, we will not post your interview transcript or audio recording online unless you give us permission to do so, in a separate agreement. At the time we send your transcript to you for review, we will also provide a consent form asking your permission to post your interview transcript and/or audio recording online. The form will provide you with different options for how, when, and with whom the materials may be shared. You will, of course, also have the option not to share the materials beyond the Duke and Adelaide researchers.

One risk of this study is that you may voluntarily disclose identifiable information that later could be requested for legal proceedings, or otherwise be used against you. Please take this into consideration when you are speaking. There may be other risks associated with your "on the record" views being made publicly available, such as having your views mischaracterized or misunderstood.

The main benefit of participating in this study is ensuring that your side of the story is properly portrayed in this history of the Bermuda Principles, which have become a model for open and collaborative research in genomics and other fields.

To help us protect the privacy of those parts of your interview that are not public, we have obtained a Certificate of Confidentiality from the U.S. National Institutes of Health. With this Certificate, we investigators cannot be forced to disclose information that may identify you, even by a court subpoena, in any U.S. federal, state, or local civil, criminal, administrative, legislative, or other proceedings. We researchers can use the Certificate to resist any demands for information that would identify you.

The Certificate cannot be used, however, to resist a demand for information from personnel of the United States Government that is used for auditing or evaluation of federally funded projects or for information that must be disclosed in order to meet the requirements of the federal Food and Drug Administration (FDA).

A Certificate of Confidentiality does not prevent you or a member of your family from voluntarily releasing information about yourself or your involvement in this research. If an insurer, employer, or other person or institution obtains your written consent to receive research information, the researchers may not use the Certificate to withhold that information.

Signature _____

Printed Name Elbert Branscomb

Date 10/12/11

If you have read this form in its entirety and agree to the interview and its terms, please sign and date above.

Contact information:

Rachel Ankeny, Ph.D. (University of Adelaide)
rachel.ankeny@adelaide.edu.au
+61-8-8303-5570
Kathryn Maxson, B.S. (Duke University)
kat.maxson@duke.edu
(919) 668-0791
Robert Cook-Deegan, MD (Duke University)
bob.cd@duke.edu
(919) 668-0790

If you have any questions about your rights as a research subject, you may contact the ***Duke University Institutional Review Board*** *at 919-684-3030 or ors-info@duke.edu.*

BCD: Just so we get it on the record, we've just finishing discussing the fact that [EBranscomb] has returned the informed consent statement. And let's go over a little bit about what's going to happen before we even proceed into the substance. Basically we will be recording this. We will send you a copy of the raw transcript. Usually it's taking about a week or less to get the audio turned into transcript. We'll send that to you and you can make whatever revisions, deletions, additions, whatever, send that back to us. And our hope is that we can … and when we send that information to you, there will also be a form that says, never release this over my dead body, release this over the dead body of my first born, one year, two years, release immediately, whatever, a whole check box. You just send that back with it and the transcript will be governed by that document, which will be separate from this informed consent thing. And our hope is to construct a bunch of archival materials that we'll post initially on our website, but we hope eventually will go into a more archival … because our grant's going to end at the beginning of 2015 and we'll hope that somebody else will pick this up and Cold Springs Harbor or Wellcome Trust or somebody will post this permanently. So the idea is to immortalize your current … it's not going to be your face. The video is not going to be archived. The audio is not going to be archived. Just the words that you say in the form that you send them back after revision.

KM: So we have a script of several questions that we're hoping to cover. Usually … we've done about five of these now … and they've pretty much gotten covered in just the context of the conversation, but I would say maybe just start out with who you are and what you do and a little bit of your history and why you were invited to the meetings in Bermuda.

EBranscomb: Well I'm [EBranscomb]. I rather stumbled into the involvement in the genome project when DeLisi started it in the DOE and anointed three laboratories to be the DOE's genome "centers". One of them was Livermore, where I was employed. And since my background exposed me to some mathematics (my formal education is in theoretical physics), I was asked to get involved in managing the informatics and data management side of the LLNL effort at that time, which was focused on cosmid-based physical mapping of chromosome 19 (under Tony Carrano). And as an indirect result of that, I got more involved in DOE and joint DOE-NIH committee level structures and via that, ultimately, to a fair degree in the grant review and advisory process for the NIH.

So that cooked along until Ari Patrinos took over BER and came to the conclusion that he needed to integrate the three efforts that had been started at DOE. That was about the time that there was a decision to speed things up dramatically in the HGP. And Ari came … I think I can paraphrase his judgments at the time … by going to few genome meetings - to the conclusion that DOE was not … having the impact he thought it should have, maybe not even having a perceptible impact, or anyway very slight. And I remember him asking me whether I thought that was due to any prejudicial treatment on the part of the DOE or was in some sense

merited, and my view was that it was essentially merited. It was not a response to prejudice. And so and he felt … especially since everything was suddenly ramping to a much more serious scale - that the only way that the DOE could stay relevant to the genome project was if he could put the three efforts together, bring them under single management and amalgamate them into a single, tightly managed effort; essentially having it exist at one site devoted to being "sequencing factory, but one that was located at none of the three national labs that were involved, and was essentially an independent facility. So that was in I think '96.

And there was a casting about of course for someone to become the director of that effort. I know it will sound like false modesty if I tell you the facts about how I came be selected to mange the new 'fused' genome effort, but it would be amusing to you to read the letter that Marv Frazier wrote for my retirement, which was read at the little retirement ceremony in 2007, in which he describes this history and says … and lays out what they were looking for in a leader. "We wanted someone who was experienced, someone who had a standing in the community" and various things like that. So everybody is sitting there expecting … they think they know what Marv is going to say … but instead the letter reads, "Unfortunately, we couldn't find anyone like that who was willing to take the job." [Laughter]. And I almost burst into tears at that point. I want to say, just because it was so honest, it was so accurate, and so "respectful" (not dishonoring either of us with falsehoods). And it's indirectly of course, flattering. And indeed that was exactly the case. They had asked … sought around in a pretty serious way to try to find someone and couldn't find anyone, and then finally David Galas, who was, as you know, Ari's predecessor and who I think you also know made his transition from theoretical physics to biology by walking into my office one night and rather quickly ending up becoming a post-doc working under me, or rather over me is the right way to think about it, and then, ultimately, he took over DOE's "OBER" office – and became my boss of bosses… and during that time Ari had worked for him. I know Ari had met me briefly, and I think, knew I had "an attitude". In any case, so he, Ari I think asked David, "Well do you think Elbert will do it?" And David said, "No, he has never been willing to do anything that had any responsibility at all". [Laughter]. And Ari said, "Well ask him anyway." And then I was asked; and I immediately said yes. But it was complicated to get that thing to actually go. It's a very hard political thing Ari had to do to get the three labs to agree relinquish their autonomy and their "inertial" genome budget, and to step together. But he prevailed.

At any rate that was of course a big phase transition for me and the project because I then became the founding director for the Joint Genome Institute, which was DOE's sequencing factory. And I believe … and that was agreed to in middle to late '96. And I think … I don't believe I was at the '96 Bermuda meeting. I'm pretty certain I wasn't. But then the next … I'm pretty sure it was the '97 meeting that was my first one. And where I was now stuffed into the seat of being, of

representing DOE's sequencing effort at that meeting. I stayed in that job until, as I said, until essentially 2000, the fall of 2000.

BCD: We're pulling up our list to see what we have on you.

KM: Over on the left.

EBranscomb: You have my time I spent in Leavenworth? [Laughter].

KM: So just search his last name and he'll be on the left side. It's a very complicated spreadsheet I've set up. We've pulled together all of the names of the participants from all the meetings and contact information we can find. So that won't have it, but hit next. There we go, '97 and '98.

EBranscomb: Yes, that's my memory, good, okay. I would have been unhappy had I completely lost my mind [Laughter], but it is open to debate. So anyway, that's an adequate answer to the question, is it not? Or more than adequate?

BCD: Yeah, so you went to the 1997 meeting, and I'm actually going to want to loop back and get a little bit more detail on the politics of establishing the Joint Genome Institute. But let's not do that right now. For right now let's focus on the Bermuda meeting. One thing, just so that we mark it, it actually would be nice if you could send us a copy of that document you just referred to.

KM: The letter.

BCD: From your retirement party.

EBranscomb: I don't have it but I may be able to dig it up. I think Marv has a copy. He and I joked about it a lot. It's absolutely exquisitely lovely. He complains about my inadequate vocabulary. He's just marvelous; it's just a beautiful construction. And sweetened because it was so accurate. At that time Chuck Shank was the director of one of the three laboratories, Berkeley laboratory, and he said, "Although I have to tell you I have never seen anyone less qualified." [Laughter]. And that too was a gift, because it was just the truth. I had in fact never led anything that you couldn't stuff into a phone booth. And it allowed me to say to him because it was our situation, "Yes, Chuck, but the exigencies of the situation are that now you're stuck with me and I'm stuck with you and we're just going to have to make the best of it." And so that's … at any rate, these were fraught times for Ari. I think it's fair to say that Ari saw that the DOE's genome program was in danger of not being visible when the project was finished and that that would be very unfortunate for DOE and for his office and everything else.

BCD: So take us back to that period when this is going on. So you're in the process of negotiating into an impossible job that you're completely unqualified for, while

this is happening for the first meeting. That meeting sounds like it was happening, had already happened or was outside your orbit. But take us back into that time, and what were people … what problems were these meetings intended to solve, what did they actually solve? And in particular of course we're going to be asking about the data sharing arrangements that emerged from the Bermuda meetings in retrospect as one of the salient outcomes. So just take us back in that time. What were people doing? What were they thinking about? Why were they holding these meetings?

EBranscomb: My understanding of the true cause and effect action here is probably poor. I saw the superficial manifestations of it. That is, the Wellcome Trust had initiated and substantiated these meetings and was funding its very rapid very large effort at Hinxton. But the nominal goals were, the general belief and understanding were, so far as I saw them, that it was extremely important to have this be organized, to be a more coherently organized effort where there were standards, reporting standards of various kinds, quality standards, participation standards, therefore, data release standards and so on.

As I mentioned in the letter I wrote to Michael Morgan some months ago, I didn't so much understand it at the time but I subsequently came to think that the dominant issue, the most interesting problem that was addressed by the Wellcome Trust coming in as heavily as they did and organizing these meetings, was that by doing that and staying on it and by various other aspects of their participation, they saved the project from being a U.S. dominated effort and essentially run and managed by the U.S. and made it be, in appearance especially, but also in substance, international. The fact that the Wellcome Trust contribution was as large as it was helped to give and was critical to giving a sort of qualitative reality to the fact that this was not just a U.S. controlled, run and managed project. And I think that was extremely important for various reasons, all sorts of reasons. It was extremely important that they came in and performed that role. And in a certain sense it was … essential features of making that work were the imposition of the data sharing, the very strict, or harshly stated at any rate, data sharing requirements and quality reporting requirements. And also a critical factor was, as has been noted and was noted by Ari, the divvy of the genome. Who was going to get to do what? And it was in the beginning an absolutely critical requirement that the various participants, European participants and Japanese participants and U.S. participants, but especially I think as I recall the European participants for understandable reasons wanted to be given certain chromosomes which they would be allowed to finish as they finished them. And so there was tremendous conflict between the desire of people to be able to get credit for having done specific chromosomes: "We sequenced chromosomes X and A and D." In the end of days, they wanted to be able to lay claim to that; and the need to get the whole thing done swiftly, which force of course became very much greater when Craig came into the game. So I would say one of the biggest tensions was not just the data sharing issue but it was this territorial issue of dividing up the genome and

who gets to do what. And in the end of days, I mean so basically of course, as you know, the policy became essentially: yes, okay, keep working on your chromosomes as fast as you can. But when we get down to it and you're not done and we have available capacity then we'll all just kind of come in and the capacity is going to be used and the genome will be sequenced as fast as possible. So there's that.

The other thing in my mind was the significance of the data sharing, the data release policies, in particular and some features of the policy that arose. In the U.S. at any rate, there came early on to be the requirement that the sequencing labs could not look at the biology; they could not look at their data they themselves were producing, basically. We were to just determine every base and submit it into the database every day and not look at it. And there was a very big issue on the U.S. side about being sure that the sequencing labs did not enjoy the scientific fruits of the sequencing efforts, of the data that they were producing, and did not allow themselves to.

BCD: So there were three or four things packed into there so let me just follow up on some of them. One is, so the Wellcome comes in and changes what could have been a U.S. dominated project into, and the word that you used was international. But of course Wellcome is kind of a substitute for U.K.

EBranscomb: Yes.

BCD: How did the other countries beyond the U.S. and the U.K. stack up and what was the ethos at that time in your recollection?

EBranscomb: Well in my recollection and my perception - with all the distortions that both introduce, the fact is that Wellcome was a surrogate of Great Britain and that Great Britain by itself was in no position and would never have done anything of this sort; it really took the Wellcome Trust Foundation to do this. They had the freedom to do it; they had the money to do it. Michael Morgan and others I think understood the significance, political and scientific, but mostly the political significance of this, how bad it would look in a certain sense if the human genome was sequenced and it was just done essentially just by the U.S. They made this very aggressive move but it did leave everyone else (apart from the U.S. and the Wellcome Trust) essentially allowed to pick up scraps. And then with a lot of bad feeling and emotions and struggle it came to be: "Yes, we'll assign this chromosome to you; you've already been working on it, we'll assign it to you. But if you haven't got it done when it needs doing and there is available capacity in the U.S. and in the Wellcome Trust, we won't just wait until you're finished." And so there was a tremendous amount of tension around that because it was these large, privileged labs, mostly the four biggies in the U.S. of various sizes of big, and then the Wellcome Trust that, due to their capacity and due to their money and so on, just controlled the game.

And the only way that this policy would have been a bad thing in my mind was if it had been just the U.S. by itself and if the Wellcome Trust had not come in. But no one else sitting around the table at Bermuda had that kind of resources or that kind of support, and so it was fraught, it was difficult for them, other European labs and the Japanese groups in particular. And for various reasons. One of them was that they … it was just a capacity issue, and moreover what that actually was, was extremely unpredictable. Capacity was ramping up furiously. The '97 meeting was very amusing because we were all obliged to say (as was I for the first time standing up in this), what are you going to do next year? And that was recorded. And for me there was a whole bunch of drama and stuff that does not reflect very well on me about that; but nonetheless, it was very fraught. It went around the table. What are you going to do? And then, what are you going to do? What are you going to do? It's another story, but since the JGI was just starting out, we didn't even have our new facility. We didn't have anything. So I had to shoot in the dark. And when it was my turn I came up with a number (20Mb) and Watson was running this part of the discussion and he just erupted in a scornful laugh, as did André Rosenthal, and he turned to Rosenthal and said, "Would you like to handle this or shall I?" And then Rosenthal said, "Well I will." And so he explained why it was absolutely ludicrous for me to be saying that the JGI could produce 20 megabases according to the standards in that year. And so anyway, it was very fraught in that setting. There was a tremendous amount of tension, but it was very…but the JGI at least had the resources, if we could use them reasonably, to stay in the game at a significant level. But this was not true of essentially any others except the other large NIH labs and the Wellcome Trust.

BCD: So, two things. One is, finish that story about you said 20 megabases, so what happened in '98 when you come back to report back after a year?

EBranscomb: To add to the story a little bit, I don't know whether you know Ray Gesteland, but he had been at Cold Spring Harbor for a while before he went to the University of Utah and he knew Watson very well. And he was an advisor to DOE and a friend of mine, and when I told him this story about Watson savaging me at this meeting, he said, "Well Jim speaks his mind." But he said, "There's a thing about Jim. If you actually do well he will acknowledge it at some point. What matters to him is not [Inaudible] but is the actual stuff." And that is actually what happened.

But anyway, so we did manage to make, as I recall, 21 megabases or very close to it. But it just about killed us and it was an extremely hard effort. But by the way, I came up with that 20 million figure by asking the heads of the three laboratories that were now being forced together in a … well coerced together into coming under my hegemony or under the JGI's hegemony, I asked them how much they could do in the following year, and I added up those numbers and it was a little over 40 million bases. So I cut it in half. But, as Rosenthal had said, no one has increased their sequencing by more than a factor of two in a year or something, and these guys (the JGI) have only put in a few million bases in the last year and

so there is no way they're going to up by even a factor of three. Moreover they're just in the middle of reorganizing and the DOE's program is, as Watson famously characterized it, sequencing by hacks.

But at any rate, so we did make our goals in that first year and that was kind of important of course. But the commitment I had to make on that occasion was critical. After the meeting I went back and actually looked at things a little bit more, and I did get very worried. This is the "bad thing" about me; I got frightened that we weren't in fact going to be able to make 20 million. And I called up Marv to tell Ari that I was very worried, and Marv said, "Ari won't talk to you." And I said, "Well, *Marv*…" And he said, "No, he won't talk to you. Twenty million is what you said; you're stuck with 20 million. We'll see you next year." And in fact Ari wouldn't budge, and he wouldn't even let me talk to him about it, even just talk. And he also said, "It's in the official 'federal' government record. It's what you've got to do." So it was … an interesting thing about Ari is his management style. So anyway, that's really just a personal aside – and about my feet of clay.

But the issue that really mattered at that time was having these sorts of feet to the fire standards, in particular how much are you going to do? And your feet are really up to the fire, not just mine, but everyone around the table. And along with the production goals being so 'harshly' monitored, the data release standards and the data quality standards were also in my mind extremely critical to holding together this (in my view), extremely fragile effort which was facing strong centrifugal and oppositional forces.

BCD: So let me capture that before we move on, just to capture that. So one function of these Bermuda meetings is to create basically, or establish from a rather inchoate coalition, an accountability structure that actually felt pretty serious and scary to those of you who were making the promises at those meetings. Is that right?

EBranscomb: Oh, yes, oh, yes. This was punishment by public humiliation, which is how we do it as primates pretty much, right? Especially for those who are not safely anointed in the big guys' club.

BCD: And who are you thinking of there?

EBranscomb: Well the large well-established sequencing operations… at the time even quite well established, and becoming rapidly larger and more-established efforts of Lander and Wash U and Baylor, with Baylor being perhaps not in such a safe position at that time I believe. But we (the JGI) had the advantage in a certain sense of being from the wrong side of the tracks. We were starting out from the wrong side of the tracks. Watson had given public talks in which he describes how the genome project got started and describes it of course in very harsh terms, the DOE's initiating the project, to quote him, was a disaster. It's a separate

discussion of what one might think of the accuracy of his indictment. I happen to think the accuracy of his indictment was pretty high. He was speaking the truth essentially. So in a certain sense we had that advantage. I had that advantage of being from the wrong side of the tracks. And so in certain sense all we had to do was make good in fairly humble terms. But the tension was high. The tension was very high in particular for the people who were not on the inside, who were not in this cabal of the dominant ones, in part because they didn't have the funds to be big enough to make substantial contributions and because they did not have government support of the right kind. And for many, the character of their government support was a critical issue. Even then it was already felt that what happened there was quite harsh and quite unfair and unnecessary even in terms of the imposition of very strict, no-compromise data release standards.

But my view is the other way around. I think that "harshness" was essential, and it was essential that it be very dichotomous and very sharp-edged and enforced pretty strictly. In part because there were all sorts of reasons and all sorts of forces that were pushing hard against this project, including political forces within the U.S. who were not at all happy with the idea that significant U.S. money was being spent to produce data that would be then just made available to anybody on the planet, people who had not paid their taxes to support it. And that's not how countries should run their affairs, as many thought. And I think all of the national interests had very strong misgivings about putting money into such big efforts and then having the fruits of those efforts just be made publicly available to anybody anywhere in the world.

BCD: And so did any of that pressure come down on you specifically at JGI? What that through Domenici or through any of the other congressional channels or executive branch channels, or how? How did you feel that?

EBranscomb: We were insulated from that and I would say the United States, I believe, or in my view of things, which is not well substantiated, is that in part because the Wellcome Trust came in so heavily…that really held the U.S.'s feet to the fire about data release. And that was a situation in which I think it would then be very hard politically for the U.S. to say, no, we're going to keep our data to ourselves and try to make money from it. But, in my view, it was the Wellcome Trust coming in that essentially backstopped that. And then with the U.S. and the Wellcome Trust together on that position having the overwhelming bulk of the resources to pursue the project that then just became the governing ethos. And in the end in my view that's what really saved the project. What really saved it in various ways, in various senses of being saved, was that miraculous data release policy. The "don't look" requirement that came on in the U.S. was sort of offensive and ridiculous, but it came about for other reasons, not having to do with the international conflict but with just the ways scientists in the U.S. felt about such data at the time. Namely, that if a lab was going to get a whole bunch of money to sequence the human genome without producing a gazillion RO1

projects a month, that that would be completely unfair and bad if they could profit scientifically as professionals from any of these data.

KM: How … sorry, keep going.

EBranscomb: No, go ahead.

KM: How was this enforced? Just at the individual meetings? Or what about after the Bermuda meetings, from '98 to 2000?

EBranscomb: It was enforced at the meetings, and also the NIH, the U.S. genome-managing group then met three times a year or something like that. I don't remember exactly what. But originally Jim, but then quickly when Francis took over, Francis would run those meetings. But the actual enforcement had to do with the extreme pressure to get data submitted and the various ways it would be known who had the clones, who was ready to sequence. But just because you were under such extreme pressure to get the data submitted and get credit for it, it was not really possible to hold data back for any significant period of time. There was always a lot of tension about that, and a good deal of discussion about so-and-so wasn't putting data in quickly, and on and on. But it was held to in effect quite closely, usefully close to the ideal. And…on top of this, the technology was accelerating so rapidly that whatever was done in the first years was almost irrelevant to the final result in every sense except the development of the technology. In a certain sense almost the whole genome was re-sequenced in the last year, much of it in the last few months. The exponential function was so steep that all of the fussing that we did in the first part was kind of mocked by what happened, except that it drove this truly incredible acceleration. As you and I have discussed, Bob, it's a whole big additional story about what was critical to making that acceleration happen and who - and what technology advances were critical, and who made those technology advances happen, and where did the money come from and so on and so on. But it is nonetheless that the rug was pulled out from under the project in a way of a magical rug that you could fly on—the rug was pulled out so rapidly that the riders could almost not stay on it. It was just like that.

KM: Magic rugs that fly. [Laughter].

BCD: So were you guys assigned chromosome 19?

EBranscomb: Well, no. When the three DOE genome centers got involved and were started by DeLisi, the individual labs picked chromosomes, and so Tony Carrano picked chromosome 19. It was small and he knew it had a bunch … it had three mapped genes on it that he was interested in. [Inaudible – laughter]. It's amazing how one thought about the genome at that time. And Los Alamos was interested in chromosome 16. They picked chromosome 16, which turned out to have incredible sequencing pathologies. And chromosome 19 turned out to be the most

gene rich per…by weight chromosome in the human genome. And Berkeley had picked chromosome 3, largely because Eddy Rubin was working on a region of that chromosome – which he declared was the most interesting region in the genome (as I recall).

At any rate, but then, when we were forced together by Ari, this was the march of Bataan, right? We were driven as prisoners of war. [Laughter]. Then it was just sequence. But at that time we were still sequencing from cosmids primarily and we needed the physical maps, and so generating the physical maps was the big deal for a long time. And also, with the law, the operational law of the U.S. public effort (which was supported by the Wellcome Trust, as I recall, who proceeded in the same way), we first of all had to sequence blind, and don't pay attention to the data, and not notice anything (even if one of the juiciest genes in the genome has just come out of your sequencing, you had to look the other way. I actually had one marvelous example of that happening, and altogether several examples)…but one was just incredible.

But at any rate, not only were we to do that that way, do it blind, but we were to sequence essentially like robots, to march down the genome sequencing every base in turn, treating each base equally and make no effort at all to pull out clones that had interesting genes in them, or give extra effort to coding regions, or anything else. It had to be biologically blind and biologically unmotivated sequencing and every base was equally precious. And part of the reason for that was the need to control the fearful appetites of the people who were doing the sequencing, that if they could in any way bias what they were doing for juiciness, then the project would fall apart. That was stated over and over and over again. There was an extremely great fear about the puritanical problem. How do you keep people disciplined? When they're digging gold how do you keep them from stealing any of the gold, and to put the gold in the company coffers instead of walking off with it? And that was a huge concern. Among certain things, it was concern about motivation, that if people really started paying attention to their sequencing or were at all motivated because of what they could get out of it in terms of biology in any sense, then you couldn't hold them to their task. You couldn't keep them working at the machines just producing the data.

And so there were several other arguments and considerations - but this was one of the most striking about the U.S.'s personality in the public effort, this devotion to blind sequencing. And Sulston and the Wellcome Trust had, in my memory, similar attitudes and biases, and so they held this as a standard. It was, "just get me the sequence and then we'll dismiss you in the end." But in an odd way, this had a kind of a good outcome. And that is partly because if you took the extreme alternative - of saying you're going to go after the interesting genes in the genome the conceit that we could have had the slightest clue as to what those really were in any sense, was and is ridiculous. Though it's still to this day I think a conceit that drives much biological research - to its own injury.

BCD: So you were going to tell us one incredible story about the juicy bits kind of dilemma. Before we dance off onto another subject …

KM: Which juicy bit was that?

BCD: Yeah, tell us a little bit more about that story.

EBranscomb: The guy's name who is important to this story has just left from my brain and so I'll have to, before I can give that story I'll have to somehow have to recover that name. He's a very famous biologist. He wrote the bible on laboratory techniques that were used for an awful long time. He's at Harvard.

BCD: Tom Maniatis?

EBranscomb: Yes. Thank you. So anyway, I was sitting in my office…good for you, that's great.

KM: He remembered something.

EBranscomb: … and the phone rang and this very soft voice said, "Hello, my name is Tom Maniatis," and he started going on a little bit and I put my hand over the phone and I said…the guy I was talking to in my office was a biologist…"There's a guy saying he's Tom Maniatis… [Inaudible-laughter]. Then I heard him say "from Harvard" … holy … it is Tom Maniatis! Anyway, Maniatis had been watching our nightly sequencing output… we were sequencing in a region that had a number of genes that were expressed in neurons that he was very interested in. They were very strange. They were the so-called cadherin genes, and there was every reason to believe they were playing some very, very sophisticated and wonderful role in wiring the brain together. And we had been sequencing blindly over the region in which these genes were, unbeknownst to us that's where they were. But our sequence had a gap in it. We had trouble with one cosmid, or we didn't have the map quite right. And he asked very softly if there was any possibility that we might be able to fill in that gap, because based on his studies he thought that gap was very important to him. And it was just such a sweet… and when he described what he knew about these genes or what he thought they were up to and what they were doing, it just blew me away – almost more than I could bear. It was so sweet and such a wonderful possibility, biological possibility, and so I said, "Well we have to fill the gap anyway. So, yes, we will try as hard as we can to get that gap filled and do it as fast as we can. But it goes up in a public database. We can't just give it to you, Tom; we have to give it to everybody." Which is another part of the deal here, and I think a very good principle, I just love it. But so he said fine. And so I said, "But I want to exact a price from you for this." And he said, "What is that?" And I said, "Well most of the people who are here doing this work are people who are trained in biology. And they would be delighted to hear about this story. They would love it to have

such an example of the advantageous utility of something that is done. And so I want, if we can do this and it really helps you, I want you to come out and give us a talk about your work." And that happened. By the way, we were also sequencing the homologous part from the mouse genome at that time, some parts, and I said, knowing the answer, I said, "Would you be interested in a mouse sequence of the same part of the mouse genome [Inaudible – laughter]?" And he said yes. So ultimately that was generated as well.

BCD: So how did Tom start this call? How did he know the data? Was he following your data set? What's the loop?

EBranscomb: He was searching through the database nightly. And that was going on during this time by lots of people. There was a tremendous amount of that going on at the time of people just combing the database every day or very often and looking for what was coming up to see if it had any…they would have a transcript that they had sequenced or something like that, or a piece of something and they were looking for a match in the genome. And they would cull the databases. So this sort of thing was happening to everyone in the business a lot. And of course it became more and more and more the format as it got more detailed and more filled in. And it was a very exciting aspect of the whole business for the sequencers, I think.

BCD: So was this a violation of this informal norm or is it okay if somebody calls you up and says, "I want you to close the gap in the following plot cosmids?" Is that okay or not?

EBranscomb: Yes, I think it was in a sense. But we were … our obligation was to close all the gaps and to march along and not leave a gappy genome. That was part of the Maynard Olson … and I'll hang it around Maynard's neck because he was perhaps the most vocal and insistent spokesman for this … but Sulston and [Inaudible] also pushed for this, and it was the NIH and DOE position - that we should not just sequence the chromosomes sort of randomly, we should march, march, march, march and try and fill in every gap that we could as we went along. And so what we were doing there was consistent with our obligations, except it was, I suppose in a certain sense, a violation that we worked extra hard to fill this particular gap. And it wasn't actually hard to do. I don't remember what it took, but it was something we could go back and we had some clones that were candidates, and went back and worked on a little bit and got the sequence. And of course we couldn't just give it to Tom. As I said, we had to put it in the public database. Now Tom knew it had significance for his problem, but not everyone would have known. But that was true just about to all of these data. You would hit on it and you would realize there was something very hot, and no one else on the planet would have any reason to think it was hot.

BCD: Is this written up anywhere? Did he ever write about it, and did he give this talk? Do you have a copy of that talk?

EBranscomb: I don't have a copy of the talk. The talk was delightful. His visit was delightful as well, I remember. It was a very lovely talk. He stayed engaged with us to get all of the DNA that he was interested in finished. At that time the Stanford group was doing much of our finishing and so they worked with him to get everything done. And I haven't followed his publications on these genes so I don't really know what he now thinks about those cadherin genes.

BCD: And when you say Stanford group, is this Cox/Myers?

EBranscomb: Well it was Myers at the time, I think, Myers group, because I think by that time Cox was, I'm not sure that I've got this right, but I'm pretty sure by the time that they were working on the Maniatis sequence that David was no longer part of their effort.

BCD: So he was off at Perlegen or Affymetrix or something? Okay. So I want to take us back to one other thing that we didn't follow up on but you alluded to in passing. So you were there in 1997, which as I recall is the year that Japan and Germany, having agreed to, been present at '96 where the data sharing norms are established, are coming back and in 1997 those norms get reestablished, and it sounds like there was a bit of wobble in whether Germany was going to abide by those rules, and Japan was, nobody was quite sure what was going on. So do you have any recollections of those debates?

EBranscomb: Yes, my recollections are qualitative, sort of a water painting and not very explicit. But my memory is that indeed the Germans could not commit to such a data release policy, and neither could the Japanese. But one of the issues in these cases, I believe was that in '97 there were two German groups there and also two Japanese groups and they were not speaking with one voice exactly.

KM: So by groups do you mean lab groups?

EBranscomb: Yes, lab groups. There was not the coherence of effort in Germany in a certain sense, the single point of control, as there was in the U.S. NIH with DOE lashed to the side was a single point of control. And things were tightly integrated within the U.S. as terms of policy and so on. But I think that was not true within Germany and not true in Japan. But at any rate, I would say what happened there was that it was not acceptable at that time to continue to equivocate on the topic. And essentially, as I recall, the upshot was, unless you can really agree to this, then you won't be invited next time.

It was kind of a who's holier than thou competition as, as primates, humans at any rate, cannot keep themselves from doing it seems. But I think the stringency was

very important and nothing short of a very strict policy like that would have held the whole thing together and kept all the counter-forces at bay. And so I was not sympathetic with the idea that André was abused. He was certainly put to a great deal of discomfort and was furious. And I think my feelings about him aren't just a response to his having savaged me so scornfully. But nonetheless I stood with those who wanted to enforce the standards and hold to them, in part so that we in the U.S. ... I think it was necessary so that we could say within the U.S., that yes, we are asking the U.S. government to support this public release, all this investment, even though it's going to be given away to the Germans and the Japanese and so on, which was extremely difficult to sell as an idea. And the Wellcome Trust/British contingent was backing this position very strongly and holding to the position that in the long run we – the major investors - would benefit from doing it this way. And as a result, I think, there was no political way to back out of it. I conjecture that it was probably critical behind the scenes, as I suspect must have been true, that Domenici was willing to support this policy. He was, I believe it is fair to say, the political force in the U.S. that essentially made this project happen.

…

EBranscomb: In my view this was a discipline that was extremely important to hold everyone to, and to do so in a pretty dichotomous and absolutistic style. I was unsympathetic with the idea of saying, well we and the Brits are going to hold ourselves to the standard, but you can come along and keep your data for six months or a year or four years and so on and that's fine, and still be viewed as good members of the human genome project. And so by the way, that's another aspect of this. What was the real coercion? What was the real source of coercion? Essentially it was only, we would not allow it to be viewed by the public that you are really fully contributing members of the human genome project. It was just a shaming coercion. But it had considerable power and it did infuriate, I think, the Japanese and the Germans most notably but not them alone. Other countries had aspirations to be able to be called participants in the human genome project. And indeed the Japanese did not express their anger in such an open and obvious way as, for example, did André.

KM: Yoshi Sakaki just e-mailed us back and we're going to interview him, and he seems very, very willing to do this. So it will be interesting to see what his take and remembrances of this are.

BCD: So, I wanted to ask another set of question so I'm taking us off. If we're finished with this part I wanted to move on to another thing. Do you have anything before I do that? Is it okay to move on?

EBranscomb: Sure.

BCD: What I wanted to capture now is, at the time it sounds like this is establishing an accountability structure and the data sharing rules are a means to that end, that probably … here I'm putting words in your mouth … maybe acquired greater significance in retrospect than they felt like they had at the time. It seems to us as we're reviewing the literature like the Bermuda rules have become something of a touchstone and an exemplar, almost an ideal, an aspirational idea of if you're going to do something that's a big deal and publicly funded you'd better do it this way and share your data. And this is the exemplar that's held forth. How appropriate do you think that is? Is that right? Is it wrong? To what extent was this context dependent and to what extent is it appropriate as a model?

EBranscomb: So I think it's appropriate as a model. It has worked, sort of insidiously and progressively, as a force to more and more and more push for data release from publicly funded research. And so I think in a certain sense, just as the scientific value of sequencing the human … of that project, of sequencing the human genome … was only in the 100th part the actual sequence data for the human genome, it was in making the first huge critical step in transforming the technology and the capability of doing biological research. That in my view is the sole reason the human genome project, but very much more than the nominal target that it had originally.

By the way, I would like to tell another story about this. At the first Cold Spring Harbor meeting that Watson held, as he puts it, NIH unable to stop the DOE just co-opted the project as the next best choice, and he held this meeting at Cold Spring Harbor. It was very packed and he was up at the front, and he was talking about … and he had been negotiating with Congress and so on and so on … and some scientist in the audience got up and said, "Dr. Watson, it makes absolutely no sense to just sequence the human genome. In order to understand the human genome we're going have to see what the genome …" and blah-blah-blah, and he got about that far and Watson said in some words that I cannot literally capture but it was in effect, "Sit down and shut up, you childish, naïve person." [Laughter]. The tone was, of course, nobody in this room misses this point. But, he said, "If I go to Congress and say, the sequencing of the human genome is just the toe part of what we've got to do …" [Inaudible - laughter]. He said, "We will be dead immediately. So we're sequencing the human genome. Got it?" Apart from the gratuitous harshness of which the earnest young man was treated, the point was valid on both sides. It was absolutely right, of course we need all these other genomes to make sense of any of them, we know that, but get real here. This already is almost more than we can … the U.S. Congress can … metabolize [Inaudible]. It was just way over the top.

And then there was a tremendous amount of opposition from within the NIH research community. As you well know about, and have written about, this was called the destroyer of all things holy and principled and proper. This was the destruction of … the insidious destruction … of hypothesis-driven science and

"inch worm" biology (and, not incidentally, the end of the control of the process by the established holders of territory) [Inaudible - laughter]. And, gloriously, it is doing quite well at all of that. But I'm amazed at how much the reality in academic biology research is still, "yes, that's nice, thank heavens you're studying all these genes, but when are you going to focus?" In academia there is this view that the way it should be done is by mud-dobbing, where you … that's how science should be done and each person should not take any sort of a 'big' step beyond their 'training' or, heaven forbid, impinge on the "territory" of others, or even evidence interest in anything very much beyond the next small incremental step away from what they did their Ph.D. thesis on. I'm sorry to get off on that.

BCD: No, that was really good. So now, going back to this as a model and aspirational guidance and all that, this became something of a template for the Fort Lauderdale meetings and all that. Were you involved in any of those subsequent iterations of data sharing?

EBranscomb: Well some things early on but not very far. I stepped down, fired myself actually, in 2000 from my JGI job and then that was essentially the end of my involvement in the committee process or managing process. I did a little bit more for a while but not much. But that idea (the "genome sciences" idea) is having in my view simply astronomical consequences in terms of empowering science – in what data are becoming freely available on the web, and in driving tools at an incredible pace for the analysis thereof: expression sets of all kinds, histone-mapping data sets of all kinds, etc. etc. well beyond just sequence data. Just exploding, and publicly available! It is having an unbelievable accelerating impact on the speed at which questions can be addressed and even more on the kinds of questions that can be answered.

The thing that makes it work, oddly in my mind, is the very thing that was so disturbing for the orthodox biological researchers about the human genome project. That is, amongst other things, it was going to bury us in data that we had idea what we would do with, and indeed it has. Scientists known to me have said they're so confused by what they get in an experiment using an expression array; they get so many genes that are coming up as affected, when some one gene is manipulated, that it's completely confusing and they can't make any sense out of it, so they don't do that any more. They make custom arrays that have only the genes they are interested in on them because they don't want to see the other stuff because they can't make a neat, tidy story out of it. And it confuses them and their students say, "Well don't you understand what's going on?"

So it is exactly that richness of exposing mechanisms that you would never, "hypothesize", that you would never guess in your hypotheses to the end of your days, if you just do a whole genome experiment of some kind, suddenly the genome is screaming at you. It hit this gene over here and this network of genes. This is where the action is. Something you'd never guess. People working e.g. on

neuroplasticity and behavior, for which they think they know a few critical genes, do an expression study and they get genes that are evidently critically involved that are not part of their discipline. They're a part of developmental biology. That's not their discipline. That belongs over in other departments. It's just confusing and confounding.

Of course the genome has contempt for nothing quite so severe as the breaking up of biology into organs and diseases and separate tissues and so on. Not the way biology works. So the genome project has had this big consequence, which is still disturbing the sleep of the established research community quite tremendously, in part because it (that is biological reality) is so overpowering, so complex. In response to a talk that was given here some time ago a very good scientist said, "So I have to get out of my field. I can't handle these kinds of data sets, I can't do this darned stuff."

BCD: So there's another layer on that that I want to make sure to touch on because your background as both a physicist and at DOE might give you an interesting perspective. And that's the sociology. One of the other debates that was going on early on of course was there are going to be big teams and nobody's going to get any credit for doing any of the work because nobody's going to know who did what, and this is big science in biology and we don't biology that way. Do you have any observations about that whole set of arguments about big science and highly organized collective efforts compared to the cottage science of molecular biology *a la* 1985?

EBranscomb: Oh, yeah, I have a lot. If I could trap somebody in a genome bar and not let them out for two hours of haranguing them about this business – there are all sorts of issues here. The work itself, the gathering the data, had to be industrialized and made factory-like, and had to be based on fixed protocols and so on. That was the big transition that made it go effectively in the early going. And now it's been sort of swept aside by the super-high throughput sequencing technologies that are coming on and profoundly changing the whole business again. Now it doesn't take a big factory to do 20 genomes for your project in a few months. And similarly the power of "re-sequencing" was something that was not well-understood or anticipated back in the time when we were sequencing the "first" human genome. So the technology is becoming 'big' and problematically big. But it is the biology itself that is "big", so there is really just no escaping 'big'.

That is, it is the nature of the problem faced by researchers in biology, the nature of biology itself in my view, that is causing its research practitioners to fumble badly in that they are not, or have not been, able to address the problem on the scale and with the character that the mechanisms that they need to understand intrinsically has. If the real action going on in some aspect of biology that matters to you involves dozens and dozens of genes linked together in complex networks (as it generally does), then the classical approaches are in general just impotent to

address that problem. And they are impotent for both technical and 'sociological' reasons. What is instead required (at least in part) is 'bigger' more cooperative efforts between lots of people with lots of technologies and competencies and lots of mathematics and lots of computer savvy. It just does not fit at all into the isolated single investigator, lone ranger, hypothesis-driven "have to know in advance what you're going to ask" sort of model. And it especially does not fit the "hypothesis testing" orthodoxy that still largely rules at least NIH funding. The genome is simply too wild, woolly, and complex for anyone to move with any reasonable speed in discovering things about it by thinking they can 'guess' a useful next incremental step hypothesis by clinging to the rock face of established "truth". The powerful alternative is just to "let the genome speak" - let it, for example, suggest to you, by any number of "fishing expedition" techniques, what genes are involved and how they are involved with each other. It will give you riches of much better 'hypotheses' than you could ever have guessed in proceeding in 'pre-genome' ways.

And so I think the genomics-enabled approach to biology is pushing very hard against entrenched attitudes, training, academic reward systems, funding mechanisms, "political" forces, and probably "human nature." Many academic scientists, for example, that work in biology are prepared to perhaps do some expression analysis, particularly if they restrict the arrays to having only genes they're "interested in" - in advance. But apart from that, going after biology the way that matches the way the genome itself manages its affairs is really disturbing to and largely incompatible with the established way of doing things and, at least as much, the established way of funding them.

Patient advocacy groups can't really get behind the actual biology underlying their disease because they're focused on a particular kind of cancer or a particular kind of tissue failure of some sort or another, some particular autoimmune disease or something, and cannot in general get behind the funding of general studies of how this whole web of mechanisms works. So I think biology is going through a very difficult metamorphosis because of this, just the way the National Institutes of Health divides up health biology is incomparable in my mind with the way biology actually works and what's actually underlying disease.

BCD: And how did you … I want to make sure that we address this credit thing also, because you're talking about the organization of the science. How do you think about the allocation of credit and how that worked out for the people like…so you were one of the top minions, right? You were one of the barons of one of these G5 sequencing centers. But how did that work out for the folks that were further down in the trenches?

EBranscomb: Um … [Audible sighing]. Oh …

BCD: We'll put down heavy breathing in the transcript there. [Laughter].

EBranscomb: The allotment of credit never maps very well in science to the creativity and the talent and ability that actually produces it. It's always quite discordant in my mind. And I think there were lots of examples of people who deserve credit who got none. And not just because it was a big lots-of-people effort, but because there was big money lying on the ground of all kinds, all senses of money. And so the powerful can see to it that the credit and recognition rebounds to them at the expense of others. But I think most people who were actually involved in this phase of the whole thing, in the sequencing of the human genome up until 2000, nonetheless found the rewards pretty intense, and continue to find them because they are producing data that is of very great value and is very, very exciting in some sense. It's just a trip.

In the Joint Genome Institute the people that we had on our three staffs when we were forced together—very many of them and overwhelmingly, those that had more advanced degrees, decamped pretty quickly. This is under my able leadership. And because we had to filter for people who were motivated and fulfilled by this kind of activity the change was quite traumatic. In our particular setting the way it came down hard right when it was coming down the hardest was, "no, we're going to work from laminated protocols. You cannot use your own little version of this protocol when we're trying to do the sequencing. You have to do exactly the one protocol that we have decided we're going to do for some months. And we're going to laminate it and you can't write notes on it and so on." So that was kind of our little joke about the discipline. But that was just awful, and insulting for most people and people shouted at me, "I'm a scientist. What I enjoy is developing my own techniques." And, yes, of course, my heart goes out to you and I hope you find a fulfilling place to do that, but this isn't it. And so I don't think I've addressed your question very well. But I got a phone call several months ago from someone who had been junior to me in this, but pretty high up in our organization, saying that this person was very upset that in the telling of the genome story his/her name never appeared. And I then quoted Ari to him. I said, "History is written by the victors, never by the vanquished." So it is those who can control this story write in the way that suits them. So, yes, that's too bad, but you don't get far losing sleep over it. And you are cursed if that's really your reason for being in the game in the first place. But I want to be explicit on the point that whereas I feel a number of people have been, in one measure or another, insufficiently credited for their role in the HGP, I am not one of them. (And, of course, "*.... the lone and level sands stretch far away.*")

BCD: So I have to disappear for another meeting that's scheduled for right now that I'm actually already a little bit late for. But this has been absolutely fabulous. But I actually want to ask, I'm going to be rude and I'm going to ask you two questions and then I'm going to disappear and leave you and Kat to finish up here. And those two questions are, I want to for sure get your perspective on Craig, what his role was as he entered the game in 1998, and what was the significance of that,

how did it interact with the Bermuda rules and compliance or noncompliance? That's number one.

And the other is to invite you to think of some of the unsung heroes and maybe put them into your transcript. If there are names, you said there were some people who didn't get credit, and this is an invitation to at least get them into the archival record via your transcript. And now I'm going to disappear and be rude, and thank you so much, and you'll get this transcript in the next couple days, or in the next week or so.

EBranscomb: Okay, thank you, Bob. So, Kat, before we launch on those topics there is another one that I would like to say some words about.

KM: Sure.

EBranscomb: And it bears on the Craig Venter situation. I think it was very important to, as essentially an implication of the data release standards and so on and the submission of the process to the control of the committee and the group, that privately funded folks who wanted to hold back the data for themselves and so on were not welcome. And I think it was extremely important to hold to that principle. If you said, we'll sequence the genome and we'll pick the top few hundred genes that we want to commercialize and we'll put the rest of the data into the public pool… I don't [think that's an idea that should have been countenanced]. So that's all.

KM: Yeah, thank you. That's extremely interesting, particularly for me, because I wasn't paying attention to this as it was unfolding like Bob was. Anyway, thank you. And on to Craig. What are your views on his involvement in this and as Celera was heating up in 1998?

EBranscomb: I did not share the view of Craig or of his efforts, and Celera's efforts, that the other directors in the public human genome project did. So I want to separate out issues of personality from just the impact factor, the effect on the project and leave the former issues aside. To put it in one very oversimplified form, I think Craig's coming in the way he did and forcing the public effort to switch gears the way that was ultimately done in order to essentially meet him in some formal sense as a tie at the finish line, that that had extremely good outcome from two different points of view. But it really did help science that he did that, in my view. That has nothing to do with my feelings about his motives or anything else, just what the consequences were. But also that, then the public effort stayed in it and continued on with holding true to its idea about what the goals were. That is, really try to get it to complete the genome – with essentially every base held to be (pretty) sacred. Get a reference genome and really try to get it as complete as practicable; that is, go way beyond what would be in any way justifiable in terms of trying to find profit-yielding genes (profit in any sense). And to maintain that

more scientifically motivated infrastructure for all the subsequent sequencing that is still ongoing. And essentially to replace, to completely replace in a certain effective sense, the combined genomes that were Celera's genome and the public genome in 2000 by versions of the reference genome that was improved and improved and improved. It was in the end that subsequent effort and doing it in the way that was blind to issues of "profit" (or even immediate research gain) but just trying to get the answer and staying true to the more public service view of what was going on, that was critical.

So to oversimplify it a lot, I think Craig's coming in, in a certain sense, saved the public effort. I know my colleagues don't think so at all. Theirs are polar opposite views from mine. But then in response to that, the public effort staying on the project and completing it by its own efforts, saved the human genome once again, and in several senses. It saved it as a scientific asset in a very critical way, and it largely saved it from being subordinated to the motives of profit from those who could manage to control it. And that has turned out to be, that and everything that has spun out of it since then: first of all just the ability to sequence more and more not just like the wind but like a tornado, has had as I said before, and is having an incredible impact. It has empowered so many other kinds of whole genome hyphened but extremely informative scientific methods that it is … it's just indescribably wonderful…and in my view none of that would have happened had this had been in the hands just of private enterprise. It really, really required the public disinterested effort in essentially every sense. In a certain sense we didn't even allow the institutes of the National Institutes of Health to say, here's the part of the genome that we want to be sequenced. It was, get the whole thing and go after the basic biology. And because the fundamental insight that we're still some huge period of time away from understanding things well enough to be able to productively direct research efforts when we're trying to understand the causation of anything that's at all complex, remains the governing truth here.

At any rate, in my view the major contributions that Craig made to the project are not just that he forced a much more aggressive "don't fuss so much over every single base, just get a rough draft first approach" – i.e. which forced the public project to be content, initially, with a whole genome that was done only in "rough draft." And just putting that pressure on and making that happen that way was an extremely useful contribution to the science of biology, human biology in the first instance, but not just that. But as importantly that the other side of his contribution had to do with various technical innovations that he was in part, often in large part, responsible for pushing into the now accepted stable of approaches invariably against scorn, ridicule and resistance.

I believe the first of these was the express sequence tag (EST) approach that he first started when he was in the intramural program at NIH. But the idea of getting that kind of sequence data, very fragmentary, very sloppy, short reads, inaccurate, "gene focused," and on and on, it was flatly rejected by the managers of the

public effort in very strong and scornful terms. None of them, I think, would now hold to that view (nor, I bet, admit to their prior posture on the question). It is universally recognized that an extremely useful tool in making sense out of any genome-sequencing project is to have the rich collection of ESTs.

There are a number of other technical approaches that Craig pushed that were scornfully dismissed, but which were ultimately universally adopted – with no thanks given to Craig. Craig wasn't the only one pushing some of these, of course, but he was in almost every case the point of the spear. BAC cloning and paired BAC-end sequencing was another one of these that turned out to be quite important in handling difficult genomes like the human genome and others. And a variety of such innovations like this, including "draft" and whole genome sequencing approaches (which became feasible when the sequencing machines themselves became much better with much more accurate and much longer reads. All of those things had a big impact and were, and ultimately became more-or-less standard procedure. But I believe all the things that Craig first pushed were strongly rejected by the managers of the public effort, until they were "silently" adopted and had in my opinion were overall critical contributions that had a very big impact. So there is a lot in this history that, in my version of reality, he is due credit for but is not given. But again, that's putting aside everything about motivation and all other *ad hominem* type issues.

KM: Right. We haven't, in any of the other interviews, we haven't really spoken about Craig too much, and we haven't talked to him yet. And he'll be one of the later people that we eventually loop around to. So just following on Bob's question, is there anyone that you'd like to get into the record here? You'd like to mention their contributions to the effort that you were involved in first of all. And second of all, is there is anyone that you think we should interview, attendees at Bermuda or otherwise, who would be really informative to this story, that you think we might not have hit on?

EBranscomb: Well I think I'd like to think about these things a little bit. So one of the issues here in this whole genome story is the overwhelming importance that technological changes made to the project. When we were first struggling with it and first planning it, to get 300 base pair reads without having paired ends or anything like that was hard, and the big slab gel machines that were being used were extremely problematic because the lanes ran into each other. It was just … and so in a certain sense it's fair in retrospect that we probably would not have really been able to put the genome together at anything like the time or cost or anything like the quality had we been stuck with those technologies for any length of time.

And so the critical thing that happened was really a series of technological advances. Different cloning procedures, from cosmids through "fosmids" through BACs. The development of paired end approaches. Etc. [Note that there was for a

time a huge push into so-called YAC cloning with huge YAC mapping projects undertaken. But the whole thing was a disastrous failure and embarrassment really and is now completely "forgotten" in the history of the genome project.]

Then there were also a large number of critical advances in sequencing technology, the development of capillary sequencing and florescent-based sequencing. Capillary sequencing, advanced polymerases, and similar improvements in the rest of the enzymology involved in sequencing, were absolutely critical advances. And also some aspects of the data management and analysis, of developing better assemblers and so on. So then the critical question is why did these advances occur and who put the money into it, who saw that it needed to be invested in, and who made it happen? And I think the history of that is illuminating, at least as I see the history, in a number of ways. It gives rise to quite a different, i.e. heretical, picture of how it all came to pass.

In my view, developing technology for this project was not something that NIH made significant contributions to – or even thought should be done in some key cases, and that not much of the credit for the critical technological advances is due to NIH. Their funding mechanisms are largely are not very well set up for technology pushing, as I think is well recognized by them. On the other hand the DOE, especially at the beginning in my view, advertised themselves, since they had the big national labs that they were going to make significant contributions to the technology. Whereas in my view they made essentially none. And so there is another question about why that was the case, assuming my picture of the matter is correct. Notwithstanding these facts, I claim, of the bulk of the key technical advances and important technical investments did come from the Department of Energy and were funded by them—just not by way of the National Labs.

And here I think the bulk of the credit for understanding what should be done and sticking with it - and seeing to it getting done - was due to Ari's office, but in particular to the fact that Ari supported the critical decisions. In particular the decisions of his subordinate, Dr. Marvin Frazier, who was managing the genome project for him and whose understanding about what kinds of investments should be made essentially defined DOE's investment efforts. For example, in the development of polymerases for sequencing, in the development of fluorescent capillary sequencers, in the development of cloning technology (especially the BAC clones and BAC end sequencing). All of was Marvin's doing and all of it played a critical role. So he is first and foremost the person who is recognized least who deserves the most credit in this history. But my version of what happened here and the kind of credit, backhanded credit at that, that I'm giving to DOE, is not a version of this history that is shared I think by any of the NIH-funded directors.

KM: Can you hear me?

EBranscomb: I can now, yes.

KM: I'm sorry I think we got cut off. Great. Well, that is so helpful, and we don't want to take up any more of your time. We've already been going on for almost two hours. But thank you so much for all of this. And we'll make a transcript. I'm going to send this right to our transcriber. And send it to you hopefully within a week. We're going up to Harvard this … the end of this week to interview George Church and Wally Gilbert for another project we're doing on DNA sequencing more broadly. And so that should be pretty interesting. But we will hopefully get this transcript back pretty soon. Thank you so much.

EBranscomb: It's been a pleasure.

KM: So nice to cyber meet you.

EBranscomb: That's very nice, thank you.

END OF RECORDING