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 Richard M. Myers

Printed Name Richard M. Myers

Date November 24, 2011

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.*

Archiving Permissions Form: "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

A short while ago, you participated in an interview with investigators engaged in a research project exploring the history and consequences of the Bermuda Principles for DNA sequence data sharing. We have prepared a transcript of your recorded interview. As indicated in the Informed Consent statement for this project, you now have the opportunity to review this transcript and make deletions and corrections.

Your transcript has been sent to you in either electronic format (via Dropbox.com or e-email communication) or hard copy format (via postal service). Please follow the instructions provided with your transcript when making any changes and when returning the document to us. These instructions are specific to the format in which you received your transcript. If you do not want to make any changes to the transcript, please let us know at the time you return this permission form to us.

In addition to the use of your interview materials in our research, we seek your permission (subject to any restrictions you impose) to place the edited, written transcript of your interview, and any related documents, on the Internet in institutionally affiliated, digital archives.

These archives may include:

- Archives affiliated with the **Institute for Genome Sciences & Policy**, Duke University.
- Archives affiliated with the **Duke University Libraries**.
- Archives affiliated with the **Genentech Center for the History of Molecular Biology and Biotechnology**, a part of the Cold Spring Harbor Laboratory (CSHL) Archives,[1] or
- Archives associated with the **Human Genome Archive** at Georgetown University.[2]

Members of the Duke University community, students, faculty and staff at other institutions, or members of the general public may access these digital archives for purposes unrelated to this research project on the Bermuda Principles. Typical research uses of interview materials include scholarly or other publications, visual presentations (i.e., powerpoint presentations), exhibits, class projects, or websites. However there may be other uses made as well, since the materials will be available to the general public. Investigative reporters and lawyers engaged in or contemplating litigation have, for example, used the Human Genome Archive at Georgetown.

Your permission to post the edited, written transcript of your interview, and any related documents, to a digital archive is completely voluntary. Unless you consent to their wider use, all materials from your interview will be available only to members of the research team affiliated with this project.

The form below provides you with different options for how, when, and with whom your interview materials will be shared. You also have the option, of course, not to share the materials beyond the Duke and Adelaide researchers. In the meantime, all digital materials are maintained in a secure, HIPPA-compliant drive at Duke University; paper materials are stored in a locked cabinet; and steps are being taken (i.e., via layers of electronic password protection of documents) to maintain the security of your materials during exchanges amongst the Bermuda research team and between researchers and interview subjects.

---

[1] The Genentech Center at Cold Spring Harbor Laboratories was established in 2006 with a gift of $2.5 million from Genentech, commemorating the 30th anniversary of the company's founding. The mission of the Genentech Center is to identify, acquire, preserve, promote, and provide centralized access to the original papers, correspondence, and research materials of the individuals and institutions that were crucial to the development of molecular biology and biotechnology.

[2] The Human Genome Archive at Georgetown University was established in 1988 under a grant from the National Science Foundation, and was long associated with the National Reference Center for Bioethics Literature and other international resources supported by the National Library of Medicine and other components of the National Institutes of Health.

PLEASE FILL OUT AND RETURN THIS FORM TO: *Center for Public Genomics, Duke University; c/o Susan Brooks; Center for Genome Ethics, Law, and Policy; 304 Research Drive, Box 90141; Durham, NC, 27708.* ***OR****: You may fax it to us at (U.S.) 1-919-668-0799.*

Interviewee Information. *Please list an address where we can contact you.*

Full name: Richard Myers Date of interview: 29 Nov 2011

Current institutional affiliation: HudsonAlpha Institute for Biotechnology

Street Address: 601 Genome Way

Phone: 256-327-0431 Email address: rmyers@hudsonalpha.org

Interviewer Information.

Full name(s): Robert Cook-Deegan; Kathryn Maxson

Affiliations(s): Duke University

I, the undersigned, have read the above, and I **AGREE** to release my interview materials, subject to any restrictions listed below:

(A) ☑ I place **no restrictions** on my interview materials.

OR

(B) ☐ My interview materials may be reviewed, used, and quoted by the researchers affiliated with the Center for Public Genomics, Duke University; *and in addition* (check all that apply):

☐ Researchers unaffiliated with the Center for Public Genomics may **read** the interview transcript and any related documents only after obtaining my permission.

☐ Researchers unaffiliated with the Center for Public Genomics may **quote** from the interview only after obtaining my permission.

☐ Researchers unaffiliated with the Center for Public Genomics **DO NOT HAVE** my permission to **read or quote** from the interview.

Posting interview materials to public digital archives: In spite of any restrictions listed above, I give permission for my interview materials to be made publicly available on the Internet by deposit in an institutionally affiliated archive:

☐ 1 year from the date of this form

☐ 5 years from the date of this form

☐ 10 years from the date of this form

☐ 25 years from the date of this form

☐ After my death

☐ Other: _____ (please specify a date or condition)

☐ NEVER: MAY NOT BE DEPOSITED IN A PUBLIC ARCHIVE

Please specify any further restrictions in the space below:

_____

Signature: Richard M. Myers Date: 6/7/12

BCD: We'll just confirm that on tape here.

KM: Yeah, we've got the informed consent.

BCD: And has [KM] already explained what we're going to do with the transcript after we're done?

RMyers: Why don't you tell me a little bit more because I think I know, but I'm not sure.

BCD: This audiotape will be transcribed fairly quickly.

KM: Very quickly, yeah.

BCD: Usually within a couple weeks. We'll send it back to you and when we send it back to you we will include a form that basically has conditions for people being able to have access to the transcript. The audio file will go on a protected drive where it's disconnected from the Internet and can't be accessed from the outside. So the only thing that will be publicly made available is the transcript. You'll have a chance to make changes in the transcript, send those back, and then we will make those available under the conditions that you stipulate with this form that we'll send you when we send you the transcript. And the other little wrinkle in the informed consent is just a little bit of explanation about the Certificate of Confidentiality. It basically allows us to resist a subpoena if anybody tries to get to you through us. So that's the purpose of doing that. I don't think it's relevant for these interviews, but it's relevant some of the other work that we're doing on patents and intellectual property and things.

KM: Yeah, and this is an informed consent specifically for this project, but also largely adapted from one that is for a larger umbrella for a set of projects that the certificate is a lot more relevant for. So, yeah, basically we'll send you an e-mail when your transcript is ready and you'll have a choice of different ways that you can get it, ranging from Dropbox to hard copy. And then you can make any changes you want and send it back to us.

And the other thing that will be included in the form—is we are trying to put together a database of information that would be on the Internet. And so individuals who are willing to make their interview information, not necessarily immediately available, but perhaps immediately available or available sometime in the future on this public repository of information for scholars and the public, will have a place to designate that. There will be an option for that too.

RMyers: And there you mean the transcript itself or … not the recording you mean?

KM: Correct. Just only the edited transcript that you have made any changes to. The audio file, once we get it, the only place it ever goes is on a secure drive here

that's HIPAA compliant and it never leaves, and on the secure server with our transcriptionist, who then deletes it after she's done transcribing. So the only thing that would ever be made public would be what you edit—and then only the written transcript, not the audio file.

RMyers: Okay.

BCD: So, [RMyers], just some senses of why you were at the meeting, what your expectations were of going to these Bermuda meetings, what the purpose was. Just some general background about both you and the context for the meetings.

RMyers: Okay, we were one of the first four NIH genome centers that started in 1990. I was at UCSF when we started it. I was the director, and David [Cox] was the co-director. We were initially set up to map human chromosome 4. It's remarkable how small our goals were. In fact, we were so conservative that we got dinged for it by the reviewers for not being bolder than that. We had developed this radiation hybrid mapping method -- David primarily, although I helped with it. And so that's why we were initially part of the Human Genome Project. We applied for one of the first grants, were awarded it, and started work in October 1990. We moved in February '93 to Stanford and that's where we did most of the work.

So from '93 to '96, when the first Bermuda meeting happened, we were having a million meetings at NIH and at various places along the way. You might recall that at the beginning there were something like 19 genome centers in the U.S. and eventually honed down to a smaller number. We started our mapping efforts starting about '95 or so, about a year before the Bermuda meeting. We started doing sequencing as well, but in a pretty light way. We were just learning how to do large-scale sequencing, at least large scale at the time. And the Bermuda meeting came, I think it was in February, I'm trying to remember; they were in the winter as I recall.

KM: It was in February, yeah, the end of February.

RMyers: Just as an aside, I had never been to Bermuda and I thought, oh boy, in the winter … not that we have hard winters at Stanford … but I was thinking we'd go into a tropical paradise. Bermuda is actually off the coast of Virginia or South Carolina. And while it wasn't wintry, still, it's not tropical. It was kind of cool there for the winter.

The purpose was exactly what you know that it was -- we had already been talking about it, not just talking but acting on -- to make data available very, very rapidly. Even from the beginning, there was this notion. I think it was partly because it was an unusual project to do such a large thing with a large number of people. Also, there were many other constituents, or users, out there. We felt that it was really important that the data are released rapidly and that we didn't get to

own it in the sense of waiting until publication before releasing it. I think that sentiment started before 1990. I remember that one reason this was discussed from the outset was because the *C. elegans* community, from Sulston, Waterston and worm groups, and Maynard Olson, really pushing this, that it was important that data is released prior to publication when you are doing something that takes a long time to do, on such a large scale and that would have such a big impact. And I think people from the very beginning, even though I don't know that any of us anticipated that we would be where we are in 2011, I think from the beginning people thought, well this will be important because we were producing a reference sequence that everybody could use. So that's where it came from, I think. And the meeting in Bermuda included the PIs or at least the heads of the projects. By then it was not very many; I think there were fewer than 30 people at these meetings. It looks like from this list it might even be only about 20 or so, not counting the NIH people. Several NIH people were there too.

The purpose was to talk about -- really to formalize -- the policy of data release even though we were already doing it before the Bermuda meetings. Of course, we didn't have the Internet in 1990, but from the beginning, data were being made available, and we wouldn't wait until publication to release it. I don't remember exactly how the release really happened for each type of data generated. The genetic maps in particular were finished early in the project. There were certainly big publications coming from Daniel Cohen and Jean Weissenbach, as well as from the CHLC (pronounced "Chelsea") group in the U.S., Jeff Murray, Geoff Duyk and Ken Buetow led that, and there were several others, part of that group, that CHLC group that did mapping. I think that while people could get to those data, it was a little different releasing mapping data than it was sequencing data because map information isn't an easy digital answer. So I don't remember how that actually went. But the principle that we're going to sequence and that we're going to put data into the public domain one way or the other, although at the beginning, it was not through the Internet. We were putting data into GenBank, and I just remember making copies of the data along the way.

But when we got to Bermuda we already knew what we wanted to do in the sense that it was a foregone conclusion that the data were going to be released rapidly. And so a lot of that was the details of, is it one week? What does it mean to release it? Do we want to release the raw reads? We talked a lot about whether we release the sequencing data when they achieve 1KB assembly, meaning when you have enough to make it more than just a raw read … so lots and lots of discussions about that. I realize I'm moving onto your other questions, but the reason we were there was to discuss those issues. I will say that the meeting turned into one of our typical regular meetings where we spent half the time talking about the science that we were doing, where people were and the progress and the lack of progress that people were making. Because in '96, the sequencing was still pretty darned hard. I think we were just barely getting able to assemble lambda-sized clones (20-25 kb) and people were hoping that we'd be able to do

this with BAC clones. By then … right about then, BACs were being made by Pieter de Jong, and that's why he was at these meetings. We were getting BAC clones of 100 to 200 KB. So while the real purpose was to develop this policy, maybe rightly so, we spent a fair amount of time talking about the science so that it fit into the policy. And I do remember…I'm just looking at '96…I do remember the international groups that were there, and I'm just trying to find the name of the fellow…it was Masahira Hattori from Japan and…what's his name…Rosen …

KM: André Rosenthal?

RMyers: Yes, André Rosenthal was at the meetings.

…

RMyers: But anyway, the contention about it [the data release policy] … and so part of my thought, and I'm sure many others must have had this as well, was that we're getting, at the time, what seemed like a lot of money to do something that was meant to be a community and a public effort and that we didn't deserve to be allowed to just hold it back like we would if we were studying globin transcription or some specific problem, hold it back until publication. It's not that you'd ever get to really hold it back completely. And so it wasn't guilt, it was that this was an obligation and it was a positive obligation. I remember talking to Maynard (Olson) about this a lot, even from the very beginning, that okay, maybe this helps if you're trying to clone a gene for disease or something, it might also help your competitors. But that's not why you're actually doing the Genome Project. If you're trying to clone a gene, it has be kind of separate from the Genome Project. And I think that's one of the things I'm absolutely the proudest of in my career just because I think it has served the world really well. It was a great model. And it's modified now somewhat but it still works pretty well.

So anyway, the meetings were interesting. They were kind of tense; I remember they were tense for me, partly because (I don't mind disclosing this) scientifically, we were the absolute smallest of the sequencing groups at that time, of all the people doing sequencing. And it was quite pathetic. We (David Cox's and my group) had only had a few hundred kilo-bases of sequence at the first or the second meeting. And in fact, it was because we were mostly mapping and we had just started sequencing. I don't think I was aggressive enough to come in and say, "Yes, I'm going to do the whole thing, give me all the money." But just historically I think it is important.

Let me tell you, this may not be important to you but it is to me. So what happened is, we had a round of funding after the '98 meeting, a round of grant applications. And by then Lander had convinced everybody, and he wasn't the only one but he was really pushing hard, that it ought to all go to one or two or three places. Meaning we don't need 19 genome centers sequencing one

chromosome each, that it needed to be global. He was right; he was dead right. Of course he didn't specify which of those three labs should be, but it was obvious what he was thinking. But what happened is that we then competed for NIH funding, I wrote a grant and it wasn't extremely ambitious but still, it was to do a fair amount of sequencing. And my score just barely didn't make it. Actually there were four grants funded. Maynard was the fourth one and my score was just a little bit lower than his. So that meant we were not going to be part of the Genome Project. I was sad about it, but I also understood. They needed to consolidate. It was definitely the right move. The lucky thing for me is, right when that happened, the DOE folks who were sequencing three human chromosomes, wanted a group to do sequence finishing for them. And so they asked me if I was interested, and we joined with them. We are still funded by them. Our genome center has sequenced something like 60 or 70 genomes of plants and other unusual animals since the human genome sequence was finished. So anyway, that's been sort of my role. This is now run by Jeremy Schmutz and Jane Grimwood, who are faculty members here at HudsonAlpha. We brought them and the genome center from Stanford when we moved here three and a half years ago.

BCD: So [RMyers], this takes me back down memory lane, and I was wondering, could you just give us a flavor for a day in the life of a sequencing center that's just getting started in doing large-scale sequencing at the time that these meetings begin to happen. What happened with the data, how much were you producing, how big did it feel, stuff like that.

RMyers: So we by then were at Stanford and our genome center was moved twice after we arrived, so that we occupied three different buildings by the time I left there. The last one we stayed in the longest and they're now trying to demolish that building, it was in such bad shape. By the way, our genome center was in the building where birth control pills were invented by Novartis (or so we were told). It later had become a Novartis plant building, doing plant research. And then we moved in. So we were there for the last 10 or 11 years I was at Stanford. So, [BCD], it was interesting because the first few years it was so new to build a big group like that, right. It just wasn't the way anybody in biology had done research. So even by '93, and even '96 and to some extent '98, a big part of the discussion … and it was true in my genome center as well as all of them … was that if we need to get this much more work done, we would need to quadruple our throughput every six months. And nobody thought it was possible to do that because the experience had been, you might be able to double your throughput in a year. But that was partly because the technology in many senses was still very crude; for instance, we were using gels to sequence. This was not even capillary sequencing quite yet. That started right about Bermuda '98, I think. And the first capillaries didn't work at all, so we spent a lot of time on that. For us, it was both mapping and sequencing; it wasn't just the sequencing. One of the things we had to learn was, going from a small lab of … well it wasn't that small, I had a lab of almost 14

people at one time when I was at UCSF, mostly graduate students, a few post-docs … and individual projects, even though they might collaborate, they were all hypothesis-driven small projects to try to understand a particular biological problem. And so the mindset of getting to the production part of this, a lot of people said, this is just going to be mindless and boring. It was never boring, ever, because we were always on the run trying to figure out how to do it. There was a huge amount of technology development. So that was number one. Number two, I think my maximum number of people that we had in the genome center was about 55 or 60, which was small compared to the 200 or 300 or so at some of the biggest centers but still seemed like a lot to me. It was a big group to manage, and a big part of what I had to learn … I don't know if you're going to talk to David Cox or not …

KM: We already talked to him.

RMyers: Okay, so David and I, by the way, are still very good friends. He's on my scientific advisory board. We're different as night and day. You could not find two more different people. The minute we met we joined at the hips and our labs became almost one lab, although not completely. We worked together for 13 years and then he moved on to Perlegen. But we still remain very good friends. But very, very different people and very different styles. And David, I think, is brilliant. He's one of my favorite people. He's really, really smart. He's just a wonderful person. But I would not put administrative organization up at the top of the list of his skills. (He would be the first to tell you that). So that's why, even though I was an assistant professor, I was just getting tenured right about the time the project started … in fact, it was literally two days after the project started when I was awarded tenure at UCSF … but it was a little unusual for an assistant professor to be the director of something. But David felt that it was pretty critical because he knew that he would have a hard time doing it, and he would rather me do that kind of thing. And it's not that he was the only brains, he was a lot of the smarts behind it, and he certainly did a lot, and we wrote a lot of papers and did a lot of things together, but the life was just so different than the sort of individual investigator-driven lab. And I will have some more things to say about me and David later too, if you'd like to hear.

But one of the things that I remember … so we went, and I wish I could remember the year … maybe Eric will remember … Eric Lander was there. But several of us went to Motorola in Schaumburg … wherever it is, some town outside of Chicago … where Motorola had its factory to make cell phones. And this was that little flip phone when they came out. This might have been about '94 or so, '95. We went to see how they did production at an assembly line. It was really amazing, they had U-shaped assembly lines and multiple sets of them. And they walked us through each of the steps. There were something like 47 little stations, and only a few of them had people at them. Many of them were robotic and there was a lot of automation. At the very end, they take the finished phone

and hold it up in a little device a foot and a half off the ground and then drop it to see if it breaks, I guess. So there was this really neat process. What we learned from that … this is relevant, I'm not completely off topic here … was, and people that when you have a process that is complicated from start to finish, if you don't do some kind of QC or quality assessment, some kind of testing along the way, then you get all the way to the end of those 47 steps and find out that the whole thing failed, maybe at Step 5. So Motorola had basically something that assessed whether each step worked or almost every step worked. This was a big lesson for us, even though I think people knew it and maybe were trying to implement it before then. It was great … not even an analogy, it was really the same principle that we adopted … which is that you do some QA along the way. For example, if you're going to sequence a clone, you first make a library. You test your library first. You do this at every step. And of course a good scientist might already know to think about those things, but the steps we were doing were extensive and complex; there were a lot of steps before you ended up at the end DNA sequence. And if you didn't do that quality control, you could end up wasting a lot of time.

There were definitely missteps along the way where … one of them, the company that made the MegaBace capillary sequencer (there were two companies making capillary sequencing machines at the beginning), but the MegaBace machine had serious problems reading one of the four bases and a lot of data was generated before people realized how big a problem this was, partly because they didn't go through all the QC and all the analysis afterwards. Anyway, that was a big part of the lesson, was learning how to work in a production environment. And [BCD], I have to tell you that, while that was all-important and I remember it really well, it seems like the dark ages compared to what we do now. Even though the technology has gotten so much better and so high through-put, we still have some of the same issues, like making sure the libraries are good and the molecular biology is working well. And just like before, we don't actually dismantle the sequencing machines and try to figure out how they work. We just rely on the companies making those and making them work well. But we have to test them all the time, and the companies don't always do that. And even with this very, very high through-put we have now, the company that is the dominant one, and hopefully there will be others so there isn't a monopoly, still has serious problems. On more than one occasion, even a brand-new machine was a lemon that had to be sent back. Stuff like that still happens.

So that was a big part of a big learning curve, at least for us, and I think it was for everybody in the field that this is a different mentality about how to produce the data. We've struggled our way and learned how to do it and of course it was with capillaries that the sequencing was able to at least get the genome done.

The first capillary machines, both the MegaBace, which was from Molecular Dynamics, and then ABI's first machines, these 3700s, worked really, really poorly. I think most of us worked with 3700s for maybe two years and basically

they all end up being doorstops because they never really worked … it was just a bad design. So then they designed the 3730 (ABI did; we still have 20 of those operating here in our genome center. So these are machines are 12, 13 years old. The guts end up needing to be replaced, but they are a very, very good workhorse for long sequencing reads). That's the only reason why we do them now, to sequence large complex genomes de novo.

BCD: So, [RMyers], take us back to … do you remember who invited you and what you thought the main purpose was of the meeting at the time?

RMyers: Yeah. I'm almost sure it was either Mark Guyer or Jane Peterson, both of them were the key leaders at the NHGRI, even though I was no longer funded by NIH by then … oh, no, no, that might have been right at the end of my NIH funding. So they were our program officers and they were essentially to help, to work with Francis on all the things that we ended up doing. So for those two meetings … I'm sure I was invited to the middle one. I'm trying to remember why I didn't go to the '97 one. There must have been a teaching conflict or something probably because I usually did my teaching that time of year. So they invited me and it was very clear that we were going to … I even think we used a phrase, we're going to finalize our data release policy because we were already as a group agreeing that rapid data release. And you know, a document came out of this. I'm assuming you have this document that never really got published and it was on the Wellcome Trust site… oh, no, it did. Sorry, sorry. I'm trying to remember if you've seen this. It's an insight outlook … yeah, yeah, okay, it's in an issue *Genome Research* in 1998.

KM: Oh, the Guyer piece.

RMyers: Yeah, so Mark Guyer must have written … oh, yeah, there, compiled by Mark Guyer. So okay, I've got that in front of me now, that's right. But I think we wanted to codify it. I'm not sure if that was the word that people used but the idea was that it becomes formal.

BCD: And so, what exactly did releasing the data constitute in 1996? So you'd produce sequence data from these highly unreliable machines. What was the flow of the information from the end of your sequencing run into GenBank or via your website or whatever?

RMyers: [BCD], I can't remember all the details, but I'll tell you what I do remember. First of all, we were, again remember, we were generating such a small amount. And we were certainly releasing it. I'm almost sure that the way that we did it then was to send it to GenBank. I don't know if it was called something else then. There were all these struggles to try to get the genome data, GDB, which never really took hold even though that was there for the longest time. I just can't remember how the database part was managed. But I do remember that that we

were supposed to release data. You might recall that these were clones. And I don't mean whole BAC clones yet. Cosmids, excuse me. The first ones we sequenced fully were cosmids, and once we really started trying to get into production mode, the main sources were cosmids. I think in the early '90s a little bit had been done with lambda clones but that was just too ridiculous (because they're smaller and much harder to grow up and produce DNA for templates). We didn't wait until we had full cosmid assemblies because the truth is that, even in '96, people were just figuring out assembly algorithms. That's about the time that Phil Green came up with PHRED and PHRAP and the CONSED module for actually looking at the assemblies. What each group was doing was sequencing small clones from the cosmids and then the BACs. We all agreed to use 3KB inserts (for sequencing templates). There was a huge fight in the early '90s up until probably '97 about whether to use only single-stranded vectors because people felt that they couldn't make good double-stranded plasmids. We were used to paired-end sequencing from double-stranded plasmids from the beginning because of a paper that Elson Chen had written where he talked about being able to piece them together based on the ends. I think that the rule was that when you'd gotten your 3KB clone completely covered, you submitted it. We didn't submit the raw reads at the beginning. That ultimately did come but we didn't submit the raw reads at first.

It turned out it was actually really useful for people to be able to get to the raw reads because the algorithms for base calling, especially on the gels, were so bad that came from ABI, were terrible. A lot of our (everybody in the field's) time was spent tracking lanes. My group did not develop any of these algorithms. But tracking lanes, figuring out how to read the sequence from squiggly, irregular lanes on gels, was difficult. Even though it was automated in the sense that sequences got read as DNA came past the laser, it still required a lot of manual handling to call the lanes properly. And that, thank goodness, was eliminated with capillaries, where you don't have the bleed-over from one lane to another. I don't know if you remember, [BCD], the lanes ran a little crooked and so you would be reading and then suddenly it looked like you jumped over to the next lane.

I don't remember how the submissions went except that it was a pain; we submitted to, and I think it was GenBank at NCBI. I don't remember if we also submitted them to GDB at the time or if GDB was still operational. GDB went on for years and then just kind of died a slow death as I recall. This was the group out of Los Alamos. Tom Marr was part of it, but there were other groups who were running that and it was meant to be this sort of centralized repository that never quite made it.

BCD: So what were the enforcement rules? Suppose you didn't submit your data, or did this ever come up as an issue among …

RMyers: Absolutely. It was peer pressure. I bet you we had five to 10 meetings a year. Being on the West Coast -- I can't tell you how many redeyes I took for the Genome Project by the time I lost my NIH funding. I was on an NHGRI study section and doing a lot of review for them as well because I was no longer in conflict. So I must have gone to Washington, I should count sometime. I probably went a hundred times over that 13-year period, maybe more.

KM: Oh, my God.

RMyers: No, I mean it was just … let me tell you another little personal thing. Actually, I'm sorry to be so … I'm a Southerner; Southerners have to tell stories. This is the way we are. You just have to understand it if you're not one yourself. But when we wrote our grant in late 1989, I think that's when we wrote it, it was reviewed in early 1990 and then the funding started on October 20th, I believe, of 1990. And so that was the official start of the Human Genome Project. They might have made it October 1 but the starting funded then. So my son, our first child, was born on October 23, and then Glen Evans had a child born right around that time, as did Eric Lander, and in January David Cox had a child. So we called them the genome babies. My son is 21 now and he's sick of me telling this story, but basically he came into being at the same time the Genome Project came to be. So we have special memories of those times, especially since it was our first child. So I'm sorry, I got off track. [BCD], what was your …

BCD: I was just asking what the enforcement mechanism was.

RMyers: So what happened was, because we saw each other so much and there were also conference calls, but because we saw each other so much and got notices all the time, it was clear, you either do this or you're not going to be part of the project. Even before Bermuda we were, the NIH-funded ones, certainly getting that pressure, as was the Sanger Center. They were part of the drive to set this policy of course. And so the truth is that, except for those two exceptions that I told you about, and a couple of other glitches along the way before that, basically we all realized we've got to do this. And so the only times that it was an issue except for the two countries that I mentioned, for the others the only times there were issues were when people were busy and they would miss a particular deadline. We set deadlines and they weren't … but I have to tell you, nobody, I don't remember anybody getting yelled at for having the wrong sentiment here about rapid release.

You just triggered a memory. There was another big contentious issue, and that held up at both Bermuda meetings as well as afterwards. In fact, significantly afterwards. And that had to do with the quality of the sequence. And you may have heard this, involving Maynard Olson in particular, but I think Waterston was one of these as well. It was hard to get good quality sequence. And we had metrics for measuring it by '98 certainly. Basically what Maynard wanted to say was that if it wasn't of the quality, it wasn't sequenced, so you can't count it. And

basically that ended up, if you really were strict about that, at the beginning it would have killed much of the sequence because we were struggling. We set this one in 10,000 error rate, which meant that you could have a one percent error rate in each direction. At first, we really realized that you had to sequence from each direction to get to that point. That would be the minimum because sometimes the error would be higher than one percent for a read. We were sequencing then about 400 base pairs, maybe getting a little bit more in each sequence read. By the end we were getting, with capillaries, more than 1,000 bp per read. The quality diminishes as you get out toward the end of the read. This was part of the discussion, even with slab gels, at the Bermuda meetings. Was Maynard at those meetings? I don't think he came to any of them, did he?

BCD: We don't have him down for any.

RMyers: No, he didn't. He certainly came to all the NIH, the ones where we all got together. It wasn't just NIH, it was NIH and Sanger. The DOE group was usually not invited to these meetings, by the way, which was an odd thing. I always thought was odd because in the end, and I'm very proud of this, in the end we produced (the Joint Genome Institute and our group at Stanford), the completed sequences of three human chromosomes, 11 percent of the total. And the quality of our sequence was one in 300,000 base pair error rate. A lot of people exceeded the 1-in-10,000 by far. So Maynard wanted to have the quality be off-scale and not to release crap. And so we had lots of arguments about that, even though I was very good buddies with Maynard and still am, he pushed us really, really hard, to the point where I just remember having the discussion that if it got too extreme then we'll never accomplish anything.

*[**RMyers: Maynard insisted that none of the sequencing groups could count sequence as their contributions unless it was almost perfect. The problem then was that we could often get really good sequence for a cosmid clone, but with a part in the middle that was really hard to finish. The argument that many made was that he was being too stringent, that we'd never release any data if we couldn't have segments -- annotated of course -- amongst the finished sequences that still had problems. In my view, I'm glad he kept this pressure on, but he also went overboard and caused a lot of anxiety, because none of us were interested in releasing bad data, but we thought we should compromise for hard parts while we figured out how to deal with them (and we ultimately did).**]*

RMyers: So, [RMyers], one other theme that you haven't really touched on but I know from past conversations you've thought about, and it's reflected in your own career path, how in a project of this scale and where data sharing is happening early and publication doesn't play the same role that it usually does, how do people know, how do people get tenure, how do they get credit for their contributions?

RMyers: Okay. We had lots and lots of discussions about that. I remember early ones not at the Bermuda meeting. Maynard Olson … sorry to keep bringing him up but he really was such a key player in this … Maynard very early pushed that we have to think about the careers of the young people we're getting involved in the Human Genome Project. They are going to have to be judged in a different way. And universities still have trouble with this, [BCD]. You may know, for what it's worth, I am adamantly opposed to the whole process of tenure. We don't have tenure here at HudsonAlpha. We have faculty, you get evaluated. But the idea that you have to start at day one and prove that you did everything by yourself is just not the way science should work, and certainly biology now, even in non-genomic biology, isn't well-served by that model. But certainly in genomics it was really, really important. And not everybody bought into this idea, and there certainly have been struggles.

But I have to tell you, we didn't run only the genome center; David and I had our labs which were actually separate but they interacted with the genome center. And the labs did human genetics problems. I continued to study transcription as well. And so that was one way that I managed it. But it took a lot of energy to run two groups like that. Again, it helped a lot that he and I joined at the hip and were partners on this because we got tremendous synergy. I mean we really, really did get synergy. David and I published twice as many papers in aggregate than we would have if we'd worked alone, I think, because you get support, you have breadth of knowledge. Even in the early days of the Genome Project, you needed a lot of different skill sets that no single lab would have. So this was something that was very much on our minds.

Personally, as I mentioned, I got tenured. I have to tell you, I'm going to admit something personally about myself that maybe is a little weird for you to hear. I started my faculty position in the beginning of 1986 at UCSF, and I had studied transcription with Tom Maniatis as a post-doc and Bob Tjian … he runs HHMI now … as a graduate student … hard-core biochemistry and molecular biology, studying DNA replication and transcription. But I got interested in human genetics when I was in Tom's lab. So as soon as I got to UCSF, Bruce Alberts, I told him I wanted to work on Huntington's Disease. And Bruce said, you ought to go meet this guy David Cox. He's a medical geneticist and does research in human genetics. So I met David and we just immediately decided we wanted to do work together. We didn't physically join our labs. He had a lab on the 15th floor, I had a lab on the 8th floor. But we had shared group meetings and shared projects. I oversaw some of his postdocs more heavily than he did, and vice versa. So in that sense we were kind of like one lab. And I just remember getting warning from some of the old-fashioned guys there, old-fashioned faculty, that, "You know, you need to make sure that you prove that you're independent." And I didn't listen to them, possibly to my peril. I was naïve. I didn't realize tenure was important. I just didn't pay attention to it at all. I honestly didn't give it a thought until my chair came and asked me for my CV and stuff like that and said

he was going to put me up for tenure … and he actually did this a little on the early side. I didn't think anything about it because I was just so busy in the lab and teaching and overseeing all sorts of things and writing grants. And then the next thing you know I get tenure. And that was very lucky because I had absolutely no anxiety about it. Sometimes ignorance is bliss. In retrospect, I'm just grateful that for some reason it didn't get crammed down my throat and I'd worried about it so much.

When I was chair of the Genetics Department at Stanford on I don't know how many different tenure committees, both inside and outside the department, I was responsible for mentoring junior faculty. And every discussion I had with them, were their serious concerns about tenure. I mean, it was terrible. And it turns out they had good reason to worry because occasionally somebody didn't get it even though I think they deserved it. So because of my personal experience, it was helpful that I understood it by the time I was chairman and really had to take care of it but I was so naïve about it at the beginning; to me, it's almost laughable. And so I think that's one reason why when we started the Genome Project … David Cox said, I want you to be the director because he thought he would be a terrible director. He traveled a lot, but he was co-director and we did everything together but divided up tasks and things. And the good thing is that even at sort of old-fashioned, hard-core UCSF, that I don't know who was on my tenure committee, I have no idea who did it. Zach Hall was my chair. I should probably ask him some day. But clearly the fact that I was running this unusual project that many people at UCSF were against at the beginning. By the way, Harold Varmus was there; he'd been a colleague. And I'll never forget Harold arguing that science was going to come to an end if we started this project. But by the time he was running NIH, he thought it was one of the best things that had ever been done. So a lot of people changed their mind once they started seeing data and realizing the value.

BCD: So, [RMyers], I think we've covered at one level or another most of the questions that we, not necessarily in the order here, but a couple of questions …

RMyers: Absolutely not in the order; stream of consciousness.

BCD: That's fine, it's actually a much more interesting way to cover stuff. But do you have any documents, can you think of any people that we should be contacting that we haven't?

RMyers: You know, I really think it's worth talking to Maynard if he will just because he played such an important role. He's kind of a (lovable) curmudgeon, as you probably know. Did you read his … oh, I can't remember what year it came out … his paper in *Journal of Molecular Biology* … it was in the 2000s, after the Venter stuff.

BCD: Yeah.

RMyers: The way he described it was absolutely the events as I remember them. So he certainly didn't make anything up. Some people think that Maynard is … he is too stringent and that it doesn't serve the purpose sometimes. But I think he is a great scientist and citizen; just need to keep in mind that he might be in a pissy mood when you talk to him. *[**RMyers: If Maynard ever reads this: Sorry Maynard! You're still my hero.**]*

BCD: He's actually going to be here, he's actually going to be a visiting professor here all next semester.

RMyers: Oh, excellent, excellent. Okay, and [BCD], you know, you probably have talked with him quite a bit, he'll have a perspective. The other one, and I'm not sure … he didn't come to any of the meetings; he wasn't invited actually, is David Botstein. And let me just give you my perspective, and again a little careful about … well this isn't off the record, I'll say. David was actually quite against the Genome Project at the beginning. He was one of those well known for arguing that this was going to destroy science. And I got to be very close with David after years of being with him, overlapping with him at Stanford. I knew him before that. But he's bombastic, he can be pushy oftentimes, and he can be dead wrong, but he changed his mind about this and so that's one of the things I really compliment David on, is that he certainly saw the value. If you talk to him he'll probably complain that he didn't get enough money. Francis messed up because he didn't give them enough money to do analysis, et cetera, et cetera. At the same time, David has some of that perspective. I would just temper it a little bit with understanding that he started out against it, became very much in favor of it. He played a role in yeast and then didn't really work on the human stuff but was involved in a lot of meetings. The NIH continued to invite him to the strategy things. I even remember one, four or five years ago that he attended. And he had a lot to offer at that meeting. And then of course David Cox, but you said you've already talked to him. I did a double-take when we were talking. I knew that Tony Carrano had died, but I didn't realize that Glen Evans had died. It's a shock. I just looked it up while we were talking. Glen was involved at the beginning and then got out of it by '98, I think he might have been or not long after that because he didn't get funded as well. It must have been '99 when I did not get my NIH funding and then switched over to DOE. And Glen didn't get funded as well. I just remember him laying off all of his people. He was at Texas then. It was kind of bad.

BCD: And you said you have some photos …

RMyers: Yeah, let me … I will e-mail. I'm just trying to see if any of them are too big. No, they're all … let me send them and if you want, I'll look up to see if I can find some better-quality ones. I've got a picture of the T-shirt that we got from the '98 meeting as well as these notes and stuff. I'll send them to both of you just in case there's a glitch here.

KM: Oh, wonderful. And you have notes, your own notes, from the meeting?

RMyers: You know, I don't think I do.

KM: Oh, okay.

RMyers: I never throw things away. So I probably do and I actually will look those up. Let me tell you, Michael Morgan was asked to get pictures and stuff because he was doing some kind of archival thing that I think he has subsequently dropped out of. I'm not sure. This was about a year ago. And Cold Spring Harbor got involved in this because they have all the archival stuff. So I dug up a whole bunch of pictures and stuff from my archives, and the way that Michael wanted me to do them was to put them on Facebook. So I opened a Facebook account.

KM: He has a Facebook group for this …

RMyers: So I annotated a lot of my slides on the Facebook thing. And I don't know if they showed up in his or not.

KM: Can I friend you on Facebook?

RMyers: Yeah, I don't use it the way like many people do. But, yeah, do that, and I don't mind you getting it. But you're just going to have to look for the folder that says Genome Project Stuff. And there are a lot of things in there, but I think one of the main reasons for looking at it is the annotation. Okay, let me just see. So I've got … let me just tell you what I'm about to send. [BCD], you're bob.cd@duke.edu, right?

BCD: Right.

RMyers: Okay. What I've got right now that's easy to send to you are these, there are two slides about where Phil Green got up to the board. I'm almost sure it was Phil. And wrote down the names of each group and how much they had accumulated by February '97 and then February '98 in terms of megabases of sequences. Sorry, that's what they proposed. And then what they actually did, it's claimed versus actual and then predicted. So the optimism factor then, this next slide, Phil shows the optimism factor was 158 megabases predicted over 68.6 actually done, meaning a 2.4 optimism factor. So this was in '98, obviously, because it was accumulating '96 and '97 results. And I just remember Phil kind of chewing us all out by saying, "You're promising all this and you're never going to be able to deliver it." So that was an interesting thing. There's a photograph of the group, which I'm sure you have. Then there's the …

KM: That's the 1996 one. Yeah, we got that.

RMyers: And then there's a whiteboard that I was trying to remember who wrote it and actually it says "human genomics sequence generated at large-scale centers." Let me send this. And it has a little bit about the policy on there. Let me just send this while we're talking.

KM: Yeah, thank you. It's John Sulston.

RMyers: Oh, I thought it was Sulston and I asked Michael Morgan if he remembered, and he said he's the one who took the photograph. That was it. Hearing you ask for notes, I knew Michael did too, but I think it will take me some time to do this. I may have to do this over Christmas, guys, because I've got a big grant due.

KM: Oh, that's fine. Anything.

RMyers: What I will do is, I use a spiral or a little notebook to take notes. It's not really a diary because it's messy but it's my reminders and everything. And I have never thrown any of these away. I still have my calendars from graduate school. But I just don't remember where they are. So I will go digging. And I am sure that I will have some things I wrote down at various meetings. They might be hard to decipher but they might also have little tidbits and some things that will be reminders. Certainly for me, I like to walk down memory lane, but I also want to get it right. I don't want to rewrite history because it sounds better this way or that way.

KM: Right, yeah. Of course.

BCD: The picture looks great. We got it.

RMyers: Oh, good, you got the photo. So I'm trying to remember the T-shirt. I can't remember why we did that. I mean why would you have … I guess … there was this notion, and every meeting started with, this is historical comment, and people used to rag on Francis all the time for overblowing this. I frankly didn't think that he overblew it. It was sometimes a little colorful. But it *was* historical. This was important. We were doing something very different. I don't think any of us thought … or, sorry, I certainly never thought that we were … what was the "Bonfire of the Vanities" phrase, something of the universe, leaders of the universe.

BCD: Masters of the universe, right.

RMyers: Masters of the universe, yeah. And I certainly didn't think that. This third one is 1998, that's right. But you can see the picture on it. It shows you we were still looking at these damned gels and trying to read them as they run past. It was really after that that we started getting capillaries.

BCD: Well [RMyers], thank you so much.

RMyers: I enjoyed it.

BCD: And we will send … this will get transcribed quickly. We'll send it back to you with the form for release and we'll remind you of a quest for any documents that might be background. And we're trying to work with Cold Spring Harbor and with Michael to make sure that this stuff gets archived.

RMyers: Okay, do remind me. And [KM], if you send, I'll friend you on Facebook and you can just look for the photograph folder. And I can't even remember what I have in there. I'll look at it myself. But I will try to find the notes, and then if I think of anything else I'll just jot you an e-mail. Something may come to mind. I think I rambled through most of what I remember.

KM: Yeah, that's so wonderful. Thank you so much.

RMyers: Good luck, guys. Hey [BCD] and [KM], both of you, I will be there for a seminar sometime. Greg Crawford is hosting me. I don't know if he has solidified it but I think it's … let me look quickly; I'm hoping you'll be there. I think it might be February 24th.

BCD: Oh, okay.

KM: At Duke?

RMyers: Yes, at Duke.

KM: Oh, great.

BCD: Okay, so we'll see you then. Yeah, I'll ask Greg what the details are and I'll try to make a point of being here. Maynard should be here by then.

RMyers: Oh, that would be great. Oh, I definitely want to see Maynard then too. I'll spend some extra time. My sister lives in Chapel Hill so I'll have another reason to come too. All right, guys, thanks, good luck. Take care.

KM: Thank you so much.

BCD: Bye.

RMyers: Bye.

END OF RECORDING