GEORGETOWN UNIVERSITY Joseph and Rose Kennedy Institute of Ethics

SUBJECT: Mark Pearson

LOCATION: Hyatt Hotel, Hilton Head, North Carolina

DATE: September 29, 1990

INTERVIEWER: Robert Cook-Deegan

Q: Ok, Mark, I think what I'll just do is go kind of right through. I may not actually hit these in the order that I wrote them up. I think the most useful thing would beto maybe detail how you first heard about the Human Genome Project and what it was defined as, the first time you heard about it.

PEARSON: I didn't hear it defined as the Human Genome Project. Through work that I'd been doing at Frederick on muscle and brain create incarnates, cDNA, and gene sequencing, I became aware of how quickly the technology was changing. An in 1983, when I moved to DuPont, where I was to head a new molecular biology group, I discovered that a couple of engineers in the Engineering and Physics Department had been working on new automated methods for DNA sequencing, which at that time involved P-32 labeled sequencing, using a thin- in window counter as a on-the-fly detection method. And it was clear that that wasn't going to work, and I told them so - particularly as I found they were spending money out of my budget. I told them that if they didn't have a better idea, I was going to can it.

And three months later, Jim Crover and Rudy Damm, the engineers, came back to me, along with a chemist, George Grainor, and said well, they actually did have an alternative and it didn't involve radioactive sequencing, which l had objected to. They had a new method based on fluorescence, and outlined to me the general principles, now, in 1990, embodied in the DuPont automated DNA sequencer, called Genesis 2000. It was clear that they had a detection principle and a discrimination principle that would allow very rapid DNA sequencing.

And as it happened, shortly after that, in middle-1984,I was contacted by Bob Gallo, with whom I'd had contact at Frederick, about the possibility of actually sequencing the virus that he believed was responsible for AIDS. His group didn't have any capability for doing this. There was an interest in AIDS, from a diagnostic perspective, but he was trying to develop out of the NCI and commercialize. And unbeknownst to me, he'd been talking to Danny Bolanezi at Duke, who had personal connections with some people at DuPont. One thing led to another, and that's why Bob called me, since I'd known him from Frederick.

And so we started actually to look at the problem of sequencing HIV (then called HTLV- 3), and did so using radioactive sequencing, and used what at the time was a relatively novel approach to do the genome sequence of this virus, and that was to actually use directed oligonucleotide priming to walk through what in those days was a big sequencing project, now is smaller, since the virus was only roughly 10 kb.

In doing that project, realizing how quickly things could actually be pushed, with a number of people using current radioactive sequencing, knowing that the rapid fluorescent sequencing was coming quickly, and having contacts with a lot of molecular biologists, I actually started, I guess in late '84, certainly early '85, to talk to people about the feasibility of actually

doing large-scale sequencing, and heard that Bob Sinsheimer had called the meeting at Santa Cruz to consider the feasibility of doing the genome, based on the development of di-deoxy sequencing, the development of rapid oligonucleotide synthesis, and the development, as Sinsheimer knew at the time, of automated methodology in Lee Hood's lab.

My other connection in this was actually through Sydney Brenner, who was a consultant to us at DuPont at the time and the person who heard about the approach to fluorescence on-the- fly sequencing that we were trying to develop. Sydney had been interested in Caenorhabditis and recognized it had a small genome that might usefully be done and therefore was part of this general debate.

So I was into the Genome Project, through all these peculiar connections, before it was a Genome Project. As we started to look at the technical feasibility of rapid sequencing, the problem became defined in part through the OTA group that you were responsible for and in part through the NRC committee that was actually asked to look at this from a serious perspective. And by that time, it was being looked at seriously. So I never heard about it as a Genome Project, but grew into it through that peculiar set of connections.

Q: You know, as I recall, the way we got your name for the OTA project was . . . I was another DuPont person from ... He was on a different OTA panel. Wasn't it a biotech panel or something? And he said you were interested in our project. I guess word must have gotten out through the networks one way or another that OTA was going to do something.

PEARSON: Prior tothat, when I wason the low-level mutation detection panel

Q: Mike Gough.

PEARSON: Mike Gough, exactly. And Mike and I knew each other through old days when he worked on P-1 virus and I worked on phage lambda, And we were definitely, at the time that program started, we were developing those technologies. Now this must have been after the AIDS sequence was done.

Q: When did that start?

PEARSON: That started in ‘84. The paper was published in January ‘85' we did it in the fall of ‘84.

Q: Was this done with P-32?

PEARSON: Yes, all done with P-32 and S-35 sequencing.

Q; You described a little bit of the technical genesis of the ideas for the fluorescence sequencing. Could you elaborate a little bit more on that? Was that something that they had been working on before, o r . . . ?

PEARSON: Well, to go back to the origins of this program, the Engineering and Physics lab at Dupont has a responsibility for contracting, research contracting, two other units in Dupont, to

solve technical problems. And two of these people, Jim Crover and Rudy Damm, one an astrophysicist and the other an engineer, had become involved in biology, developing a super- sensitive detector for an amino acid sequencer, and their technical scientific background and interest was really in signal processing. Through that connection with the biologists who were focused on sequencing problems, now on proteins, and focused on amino acid technology, they learned of the need for the rapid increase in those days of an interest in DNA sequencing.

And Dupont, surprisingly enough, not having an army of people, was being very amenable to instrumentation and automation for repeatable tests. Scientists in Dupont recognized that they might be able to get money to develop an instrument, but would not be able to get money to buy the people to actually do the work. So this was actually an end-around to try and acquire sequence information. This was driven by need, to begin with, and not by the desire to build a machine.

It happened that these two guys were in the right place at the right time. And they happened, by good luck, to meet up with George Trainor, who was trained at Columbia as a carbohydrate biochemist and he didn't know much about nucleic acid chemistry. But George knew that one of the chemists in Dupont in the middle '70s, had developed a set of chemical reactions that made it possible to tune organic dyes, by hanging appropriate pendant groups on the dyes, shifting the spectral properties of the dye. And George recognized that the four-base discrimination devolved into a four-dye problem, and required for its appropriate solution structurally very similar dyes, so the chemistry basically could be developed and would be the same thing. So that was the origin of that particular triumvirate of the two engineers, or the engineer/astrophysicist with the chemist.

The idea, then, was to build a machine that would be used in-house and would be nothing more than a special-purpose, for-our-own-use machine. And it wasn't until much later (and after a lot of skepticism and whatnot that's subject to another story some day) that we ever decided to try and make this into a commercial instrument.

So the idea behind developing rapid DNA sequencing methods was there in 1983, was developed through '84, '85, '86, and basically that project went for three years with no real results before it all started to gel. Because new chemistry had to be developed, both on the dye front and on the nucleotide linking front - and that turned out to be the key element - and also on the instrumentation laser excitation dual-filter signal discrimination side.

Q: Getting back to the Genome Project. In your view, what was the definition of the Genome Project in 1985, and what is it in 1990, and how has it changed?

PEARSON: Well, in 1985, at least as I saw it, it was primarily the development of a total genome sequence. That was a very clear-cut objective, and it was, at that time, not determined as to what organism it ought to be done in. It was clear that the smaller the genome, the less the work, potentially the greater the information. On the other hand, in 1985 we were on a downslope historically in the real intellectual interest in understanding microbial genetics, molecular biology, physiology. Really, that was an era that I see as having its peak in the '60s and '70s. And the organisms that people were becoming more interested in had bigger genomes, they were: mice, men, and com.

The original idea, as I learned of the project, was really just a macro-sequencing project.

And as one talked about it, it became very clear that there was a horrible problem, on the front end, of actually isolating individual clones. We all knew, in 1985, that it was possible to make libraries. We also knew that there were problems in propagating these libraries. And therefore whether or not one could have complete coverage, even of a complete genome, in individual clones propagated through E. coli or some other host, wasn't clear that, in fact, would work in the human(?). People weren't really aware of the polymerase chain reaction approach to sequencing and DNA analysis. We didn't know if any of the instrumentation would really work. In my mind, everything was focused really on cloning and sample preparation and sequencing.

Subsequently, in 1990 (if we just jump ahead), the importance of mapping has become much more apparent. And the congruence between mapping, to identify disease loci, to acquire useful information, and looking at individual variation in the human population and associations of genome sequence variation with disease loci, become much more apparent. Also the congruence between limited micro-sequencing and the generation of a physical map, through what we now call the STS map, clearly identifies local regions of sequence that are useful in a mapping context, but which are also handles into the ordering of clones and the generation of anchor points for a complete genome sequence.

At the same time, in 1990 the interest is in trying to do complete organisms of a variety of a species, and in particular one project I'm associated with in determining the physical map of Arabidopsis, is something that in 1985 wasn't appreciated as a useful model system. Between 1985 and 1990 it has become quite clear that will be the paradigm in the plant world for a total genome sequence. While the effort in Arabidopsis has been slow to start up, it's coming on-line now, with people who were previously trained basically in E. coli and other microbial systems.

Also, I guess, in 1990 the project is different in that we have much better computer tools, and the possibility of truly automated clone identification and sequencing, in some cases involving PCR itself without recourse to cloning, seems to me a lot more tractable.

I'd also say that as of May 1990 there were a variety of different methods that, I think, will converge on generating a good physical map. From my point of view, that was not so clear until this year. I'm not saying that it's done and that there won't be problems, but there are several independent routes, all of which look like they will give a high resolution physical map. And therefore the challenge of identifying, again in the human case, individual disease loci, or in the plant world, genes determining important agronomic traits, is, I think, the biological face of this whole problem.

And, I guess, from the perspective of 1990, there is a much greater appreciation of the ability of this project to identify multilocus characteristics, whether we're now talking about disease states or more complex traits in all organisms. And that's something that I don't think has yet really seeped into the consciousness of the scientific community.

The rapid success in isolating the Duchenne muscular dystrophy gene, the cystic fibrosis gene, and now the neurofibromatosis gene has made people think: "Gee, we're going to find a gene for each one of these diseases." More complex diseases, such as a number of the auto- immune diseases, where there are now known to be multiple loci, some of them involving immune response genes, I think it's just the tip of the iceberg.

So the Genome Project, as such, now has a set of ramifications in 1990 that I don't think it did, at least for me, in 1985.

Q: From the perspective of a company that spans chemicals, commodity chemicals, specialty chemicals, and also pharmaceuticals, how do you think the role of the Genome Project, including all the components: the mapping, sequencing, and technology development, fits into that?

PEARSON: Well, I can give you two interesting world views on that, I think. Let me give you the 1989 version and then the 1990 version, because, as you know, Dupont has gone through a major change in its strategic and organizational focus. You may not know, but this week the company was reorganized in the most significant way it has been for a hundred years.

But let me start out with the 1989 version. In 1989 there were basically three departments, or interest groups, if you will, in Dupont interested in the Genome Project. One is the Medical Products Department, made up of three divisions: a diagnostics division, a biotechnology systems division, and a pharmaceuticals division. Another is the Central Research and Development Department, in which I reside. And the third one was the Agricultural Products Department.

Let me start with Agricultural Products first, because it's almost the simplest, in retrospect, to review - the simplest set of interests to deal with. Ag Products has come to realize that genetics and chemistry can work together to modify crop plants and ultimately influence yield, both the quality and quantity of agronomically valuable traits in plants.

This is actually a new view, because the Ag Products Department has long held the view that chemicals represent the only route to the acquisition of value and use in the agricultural world. But now the Ag Products Department recognizes that it's actually in the food business,

and that that business ultimately goes through properties such as the processability of, for example, soy beans, the value of oil in soy beans, et cetera, et cetera.

So their world view has changed, and with it, they've come to recognize that genetics is important. And, from that perspective, we have both a crop protection chemicals program and a crop genetic modification program at Dupont.

The Genome Project was initially seen as something important only to the medical world. But now, through Arabidopsis, as I've mentioned, it suddenly seemed to have enormous agronomic importance to Dupont's future business.

When I started the Arabidopsis program in Central Research, the Ag Products people tried their best to discourage that development, feeling that this was not a crop plant and why waste your time with a model system. Now that they've realized that things might be a hundred times easier in Arabidopsis, and that important genes in Arabidopsis, identified and studied there (an example being that you can get in a small incubator the number of Arabidopsis equal to all the com planted in Iowa and can study this over a one-month period), is something that has come to impress the Ag Products people.

The real point being that genes identified in Arabidopsis as having potential importance can in fact be used to probe and fish out the relevant genes in, say, soy bean or corn. And hence, then, a genetic modification and transformation in soy bean and com might be greatly speeded up by knowing this information in Arabidopsis.

So, from the perspective of the Ag Products world, the Genome Project has been seen to provide information that might be commercially exploited, and particularly so if one is dealing with multigenic traits, because of the work that Botstein and Lander and Kingsley have done in

tomato, identifying five genes that are involved in the water content, or the solids content, of tomatoes. The muligenic character of quantitative traits in plants has become instantly appreciated by the Ag guys, where that still is not appreciated by people in the medical area.

Therefore, to just wrap it up with Ag Products, there is a belief that this is still fairly esoteric academic sork, but it might actually be relevant to their business, and they'd like to see a lot more of it go on, quickly.

To that end, the Ag Products people have developed a new mapping method, actually, which has allowed very rapid generation of maps of com, soy bean, and Arabidopsis, now in the space of man-months instead of years, to get high resolution linkage and physical maps. You can use probes developed by this RAPD method to identify linkage differences between plants, and that methodology generates probes that can be used to identify physical clones that are derived from the region. So it's a very powerful general method. (Which works also, by the way, in man and microorganisms.) And the development of that method has relied on automation, that is, robotics, for the preparation of large numbers of probes, automation in the oligonucleotide synthetic side, and the use of gene amplification methods to generate the appropriate patterns that can be genetically mapped.

I'll skip through Central Research, the department I'm in, on the way to Medical Products, which is the interesting example. In Central Research, we've been interested in genetics and the application of genetic methodology and technologies to the plant world, the microbial world, and human health. Given the development of the Genesis Program, and also Nat Sternberg's large DNA P-1 vectors for physical mapping, gene isolation methods, and our long-term interest in genetic recombination mechanisms, DNA repair mechanisms, and from the

cancer radiation perspectives, and ultimately in gene therapy; the genome program represents a part of our overall interest in understanding basically how organisms work, from an analytic point of view. Our long-term interest is actually in the synthetic side - being able to modify genes and behavior. So the genome project has a natural affinity and home in the interests of people in Central Research, interestingly enough, particularly on the informatic side. (We'll come back to that later, perhaps.)

The really interesting case in Dupont, however, is Medical Products. The Pharmaceutical groups have traditionally said that the Genome Project is interesting but has nothing to do with their business. (And I'll come back to that.) The people who have said that it's crucial to their business are people in Diagnostics and in Biotech Systems. And the strategy in Dupont has been basically to take an instrument developed in Engineering, Physics, and Central Research, the Genesis machine, and commercialize that through Biotech Systems. And they have seen, therefore, reagents and instrumentation as being, in their term, a media product that is relevant to the product mix of Biotechnology Systems. The Diagnostics people have assumed that these reagents and instruments will be used to acquire diagnostic information and could be combined into a non-research, much more closed system equivalent to the automated clinical analyzer that Dupont's Diagnostics Division makes, to build a next-generation capability in diagnostics.

Interestingly enough, in late 1989 the Diagnostics business made a strategic decision that it was going to concentrate not on developing new technology and new businesses but rather was going to concentrate on its "core position" in automated clinical analyzers and clinical diagnostics. They sold off their blood processing, AIDS detection business to Johnson &

Johnson, and have stopped development of DNA-based diagnostics.

The Biotechnology Systems group, selling Genesis and reagents, have therefore had what might have been a large internal customer removed from under them. And they have found, as have Applied Biosystems, Inc., our competitor on the automated DNA sequencing side, that the market is very soft for automated sequencing. In part because early instruments didn't deliver on the promise that people expected, and now, ironically, when the machines are totally functional and can deliver, (depending on who's running them and what the circumstances), twenty to forty unfinished raw kilobases of DNA sequence per day. The rest of the world hasn't quite caught up with that, so we've got an anomaly where the business basically is on a downslope but the performance is on an upslope.

Whatever all of these economic conditions indicate, the fact is that Biotechnology Systems group has slowly recognized that a single product in this game does not a business make. And therefore they're in the midst of thinking through carefully what their role is, not just in producing automated DNA sequencers at the present time, but also oligonucleotide synthesizers, et cetera. And I think the one thing that is clear is that that group, Biotech Systems, will continue to supply reagents, ranging from fluorescent nucleotides through a variety of enzymes ... for researchers, but I'm not sure where the instrument business, as such, is going.

Now let me return to Pharmaceuticals, because that's the interesting 1990s example. Almost exactly two months ago it was announced that Dupont Pharmaceuticals would be split from Dupont and a new company would be created in partnership with Merck to form Dupont- Merck Pharmaceuticals. The formation of this new company has brought in some new leadership into the Pharmaceuticals group, and there is now suddenly an appreciation of the value

of the Genome Project in identifying disease targets that are required for therapeutic drug design and the development of the next generation of pharmaceutical agents.

It has also made the Pharmaceuticals people recognize that perhaps the responsiveness of individuals in the population, particularly those undergoing early clinical trials, might actually be monitored and properly assessed by referring to those people's genotypes as well as their drug response. So we're talking now about the recognition that genotype can determine drug responsiveness. [End Side A]

We've been talking about Pharmaceuticals starting now to recognize that perhaps the Genome Project is relevant to their future business and patient (i.e., customer) identification.

I don't know where that discussion is going to go, and I for one am trying to push it very hard because I believe that the Genome Project is going to lead to more accurate diagnosis of patient populations who will be targets for a particular drug cycle. I also believe that the Genome Project is going to identify gene products whose proteins will be targets for drug design. And ultimately I believe this is the approach that will lead to new therapeutics at the nucleic-acids-as- therapeutics level, involving both antigens and other forms of nucleic acid therapeutic itself, and ultimately gene therapy.

So right now this is all a big debate going on while the new company is being formed, and I don't know how it's going to shake out. I see Dupont as a large company, as a microcosm of the issues, problems, and prospects for the Genome Project.

The one area that I've left out is Dupont's interest in the environment. It has become clear that environmental concerns will be a major issue certainly through the '90s and beyond. The chemical industry views these problems as primarily an engineering and chemistry problem,

and is slowly recognizing that there is a huge biological component.

I see the Genome Project, actually, as it's applied to model organisms and as we acquire the technology for analyzing microbial populations, as a particularly useful diagnostic tool (and possibly prognostic tool) for modifying the environment through bio-engineering, in its grandest and largest sense.

What that means in terms of cleaning up oil spills or toxic waste dumps, et cetera, remains to be developed. But I think that understanding the genes, for example, involved in the hydrolysis of, for example, cyanide, will offer opportunities in the environmental area that could have a major impact on society in this decade as well as the next.

Q: What do you sense about the Genome Project as a whole? Do you thing it's a big deal, and if so, in what sense?

PEARSON: I think it's a big deal at a couple of levels. As I've already indicated, I think it's going to shape our society in the future in a large number of very important and fundamental ways.

It's not a big deal in terms of the relatively relentless advance of science and technology. And it doesn't consume a huge amount of money at this point in time. Its potential prospects, I'm thinking now of the concern at the ethical level by not just the scientific community but by the lay public, sometimes makes it loom larger than it really is.

Yeah, it's a big deal, but it's interesting because the project itself, in terms of the acquisition of the information, is a relatively short-term, closed-ended project, as I see it.

If you look at any new product, it takes about twenty years from conception to the time

that product is actually being used on any kind of large scale. I doesn't matter whether you're talking about a new fiber, a new pharmaceutical, or a new approach to electronic devices.

And I see this project as being really a twenty-year project. It will have grown, will have matured, and I think will be dead as a project per se. It will fade back into the technology base, from my perspective, through a twenty-year time frame.

But I also believe that the information that's gotten now will be the information that's being used five hundred years from now. And there are not many human activities that I think one could point to and say: well, five hundred years from now people will probably still be doing this.

Q: I don't want to beg the question, but actually I'm kind of curious *about your feelings about it. You have discussed a lot of the technical aspects of the Genome Project. I'm kind of curious, from your own perspective, what was the process, what was the most important thing in getting it off the ground? To the extent that it's launched, do you think it's launched, and if so, what do you think were the major events, the things where people . . . ?*

PEARSON: Well, I think, like all these things, a few key personalities were the major drivers. I have a highly biased point of view, of course, but I think the success that Lee Hood demonstrated in the automation side of it. People tend to lose sight of this, but coming from a laboratory seriously interested in auto-immunity and biological mechanisms, as opposed to just from an instrument-develop, engineering point of view, I think that was important. And Lee, of course, has been an articulate spokesman for this for a long time.

I think Sydney Brenner played a key role in pushing forward. And this was a logical extension of the work that he initiated on mapping the nervous system of the nematode. To map the genome was in a sense a return to his roots as a molecular geneticist and one of the founding fathers in molecular biology.

Jim Watson obviously has played a key role, but a relatively belated role, from my perspective. He wasn't one of the key pushers early on and I think was somewhat negative about it early on.

Jim Wyngaarden I think was an early anchor on the whole project and didn't seem to have a fix, from my perspective, on what the project was about or why it was important. And, surprisingly, I think Ruth Kirschstein helped him to see that it was important. I think the Reston meeting, for me, represented a situation in which Wyngaarden basically got turned around and came to recognize that there were a large number of good reasons for pursuing the project, and that it was actually technically feasible and had medical relevance and fit within the purview of NIH, and I think he saw its special flavor. Also, of course, he came to recognize the political expediency of taking this as a new project to Congress to try to get "new money" applied to biomedical problems.

The irony is, of course, that a reluctant Jim Watson became very much the key advocate of this, supported by Wyngaarden in the end, and what seemed to me to be some bad feelings on Ruth Kirschstein's part, when in many respects she was kind of the mother of the project, in the federal bureaucracy, in a very interesting way early on.

I think you played a key role, and have done not just the OTA activity, which I think was in many respects redundant with the NRC committee, but it had the aura of putting a larger **stamp**

on it than just the National Academy. And your connections with Congress gave it a national patina of a sort that Academy, NRC committees in and of themselves just wouldn't have gotten.

There are a lot of technical players here. Because underlying all these discussions about whether it was a good idea of a bad ides, doable or not doable, was relentless, steady progress in the technology, always pointing to: This would be easier, it would be done sooner, it would be done cheaper, and it was more interesting than any of us originally assumed.

Q: I'm curious about your perspective on Ruth Kirschstein. What was her role?

PEARSON: Well, Ruth, as the director of General Medical Sciences and a strong proponent of genetics in the NIH, through GMS, for certainly the decade of the '70s and '80s, seemed to me to be the only person at that level in the National Institutes of Health who actually understood that genetics was important. I say that not to deprecate the opinions of other Institute directors, but my own personal view is that none of them had the same feel or understanding of the importance of human genetics and the relevance of microbial genetics and model organism genetics to the human condition that Ruth did. Partly that goes back to her old work on polio, partly it's because she was associated with those scientists who were pushing at the forefront of this, and so it was a legitimate point of view and a natural one for her to hold.

I fault her, in part, however, for coming back then and trying to play up (as part of the jockeying amongst the Institutes) how much of this work was actually being supported in the Institutes, suggesting early on that there was actually more total activity in a dollar amount than I think any of us actually believed. And this is, unfortunately, part of the NIH politics. It's like the

NCI turning around, arguing how much basic science was supporting them, when in fact that's a relatively small fraction of their total budget.

Nevertheless, Ruth I think did play a key role in getting this program into the federal bureaucracy, into the one agency that has broad national responsibility, that wasn't going to internalize the whole program.

And I contrast that with the Department of Energy, which has had, through the World War II experience, historical interest in genetics and mutational basis of a variety of diseases, all stemming, obviously, from the bombing of Hiroshima and Nagasaki and radiation damage. But that radio-biology approach, if you will, or culture within the DOE, does not admit the larger understanding of genetics and variation as seen in a natural population not induced by mutation.

Early on, I know there was great concern as to whether or not his program should be run out of DOE or run out of NIH - DOE obviously being interested in trying to grab all the funds as an alternative method of keeping alive the mission of the National Labs at a time when its expertise and total activity in this area was actually minimal.

Their argument that they understood systems management and Big Science, as events have shown, in fact has an element of truth to it. And in many respects I think Charles DeLisi's early models were, in concept, right, but actually not all that well executed and not backed up by the cadre of expertise and capability that the NIH community had.

The NIH community, on the other hand, has thought in much smaller, more individual, creative units, instead of large multi-disciplinary research teams. And therein lies the strength of both institutions and also their weaknesses in dealing with this project.

So it's interesting that, today at least, both NIH and DOE are trying to play to their

particular strengths in this project, and frankly both are having troubles because they don't have the strengths of the other group.

Q: Where do you see things going fifteen years down the road with the Genome Project? You said you think it's going to be a closed-ended. . .

PEARSON: I think that the Genome Project will become a set of data bases. Useful data bases, increasingly richly annotated. Used by scientists, medical scientists in particular, to understand how people work, and practicing M.D.s to actually understand the molecular basis of the genotypes of patients that walk in their front door. So I see the program basically as having a service element on the medical side and being basically a standard reference work on the scientific side.

Clearly, the project, as such, could be continued ad infinitum to sequence the complete genomes of all living organisms, but it's not clear to me that that has any great merit to it.

I think we'll start to learn an enormous amount once we get the first few model organisms actually done. I'm really looking forward to seeing the complete sequence of E. coli - as I actually already have in the case of lambda. As I worked in lambda in the late '60s and early '70s, the complete sequence of that genome identified for me several proteins that I'd been working on, characterized only as spots on gels. Proteins, interestingly enough, whose function is not yet known, even though we've had the sequence of the lambda genome for, for the sake of argument, fifteen years.

So the project, from my perspective, will become much more oriented towards biological

understanding and will fade into the mists of being a real project. Just like we no longer, in physics, talk about the ether, we won't talk about the Genome Project, either, in a hundred years.

Q: What do you think was your main contribution? What was your role in all this?

PEARSON: From my perspective, rightly or wrongly, I think I've seen the connections at both the basic and applied level through the project, and have been fortunate to have been included in the discussions, in a sense as a representative of industry, even though I'm an ersatz representative, really having spent most of my career in academic science.

I think my own personal contributions at the scientific level actually happened with the HIV sequence and the sequences early on. Where I see them going in the future (and this is something I will talk about at this meeting tomorrow, actually) is in the rational design of new drugs and in the development of alternative therapies, which is really the focus of much of our current work. Including, for example, the design of DNA sequence-specific binding proteins, which I believe will represent new therapies, new pharmaceuticals in the next century.

So my involvement in the program has been, I think, a bit of an interpreter and an advocate within industry, on the one hand, for this project and for the value of genetics as a whole. And on the academic side, I think I've become the token industrialist.

And I would also say that, actually, from my personal perspective, this had been an enormously satisfying experience, because I was a professor of medical genetics for ten years at the University of Toronto, and having worked on the development of a lot of the technology and concepts, in my work in phage lambda in the '60s and '70s. Man is just lambda, writ large, and

now its being done - in my lifetime.

And I guess the other thing I would add, at a very personal level, is that I'm interested in the Genome Project from the perspective of human disease. Both my grandfathers died from colon cancer, my mother has diverticulitis as a predisposing condition. I would like to know what my genetic status is, in that particular context, and would like to know what my kids' might be. To really personalize it, my father-in-law died a couple of years ago of Alzheimer's; I saw what that disease represented. And while it's not even appropriately recognized as to whether or not it's a genetic disease, in the real molecular sense of the word, that actually is of personal interest to me. And I see this project as having an impact on my kids' ability to understand their own gene-type, in molecular terms, and actually do something about it. And I think that my grandchildren will actually have the chance to benefit, from a therapeutic perspective, from the work that we're all doing now.

So, to me, this is one of the great pluses of being a scientist and being part of this turn of molecular medicine, if you will, that can have impact on a huge number of people.

A contrast, again at a personal level, is my brother. He is a surgeon, and he gets immediate, personal satisfaction operating on somebody but can affect only a relatively small number of people. I don't get that personal satisfaction, but I think the potential here for these activities and the sum total of all the work we're doing has an enormous potential impact on all of society and future generations. It's a very satisfying thing to do. It's kind of corny, but it's very much a driver.