DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

ADVISORY COMMITTEE FOR PHARMACEUTICAL SCIENCE

PHARMACOLOGY TOXICOLOGY SUBCOMMITTEE

Tuesday, June 10, 2003

8:30 a.m.

CDER Advisory Committee Conference Room
5630 Fishers Lane
Rockville, Maryland

PARTICIPANTS

Meryl H. Karol, Ph.D., Chair
Kimberly Littleton Topper, M.S., Acting Executive Secretary

Subcommittee Members:

Andrew Brooks, Ph.D.
Jay Goodman, Ph.D.
Jerry Hardisty, D.V.M.
Michael D. Waters, Ph.D.
Tim Zacharewski, Ph.D.

Guest Speakers:

William D. Pennie, Ph.D.
Kurt Jarnigan, Ph.D.
John Quackenbush, Ph.D.
William B. Mattes, Ph.D., DABT
Krishna Ghosh, Ph.D.

FDA Staff:

David Jacobson-Kram, Ph.D., DABT
John Leighton, Ph.D.
Frank Sistare, Ph.D.
Helen N. Winkle
Janet Woodcock, M.D.

C O N T E N T S

P R O C E E D I N G S

Call to Order

DR. KAROL: Good morning, everybody. I would like to call the meeting to order. My name is Meryl Karol. I am from the University of Pittsburgh and, since many of us are new to the committee and the subcommittee, I would like to go around the room and have everyone briefly introduce themselves with their name and their affiliation. We will start over there.

DR. LEIGHTON: My name is John Leighton. I am a supervisory pharmacologist in the Division of Oncology Drug Products. I am also the Associate Director for Pharmacology for the Office of ODE-3. I am also the co-chair with Frank for the nonclinical pharmacogenomics subcommittee.

DR. SISTARE: I am Frank Sistare, with the Office of Testing and Research in the Center for Drug Evaluation and Research at the FDA.

DR. GOODMAN: Jay Goodman, Michigan State University, Department of Pharmacology and Toxicology.

DR. HARDISTY: Jerry Hardisty, from Experimental Pathology Laboratories. I am a veterinary pathologist.

DR. KAROL: As I said, I am Meryl Karol, from the University of Pittsburgh, Department of Environmental and Occupational Health.

DR. WATERS: Mike Waters, Assistant Director for Database Development, National Center for Toxicogenomics, NIEHS.

DR. ZACHAREWSKI: I am Tim Zacharewski. I am in the Department of Biochemistry and Molecular Biology in the National Food Safety and Toxicology Center at Michigan State University.

DR. WOODCOCK: I am Janet Woodcock. I am the Director of the Center for Drugs at the FDA.

DR. JACOBSON-KRAM: I am David Jacobson-Kram. I am the Associate Director for Pharm/Tox in the Office of New Drugs in CDER.

DR. WINKLE: I am Helen Winkle. I am the Director, Office of Pharmaceutical Science in CDER.

DR. KAROL: Thank you very much. Now we will have Kimberly tell us about the conflict of interest.

Conflict of Interest

MS. TOPPER: The following announcement addresses the issue of conflict of interest with respect to this meeting and is made a part of the record to preclude even the appearance of such at

the meeting.

The topics of this meeting are issues of broad applicability. Unlike issues before a committee in which a particular product is discussed, issues of broader applicability involve many industrial sponsors and academic institutions.

All special government employees have been screened for their financial interests as they may apply to the general topics at hand. Because they have reported interests in pharmaceutical companies, the Food and Drug Administration has granted general matters waivers to the following SGEs which permits them to participate in these discussions: Dr. Meryl H. Karol, Dr. Jerry F. Hardisty, Dr. Michael Waters.

A copy of the waiver statements may be obtained by submitting a written request to the Agency's Freedom of Information Office, Room 12A-30 of the Parklawn Building.

In addition, Drs. Andrew Brooks, Jay Goodman and Timothy Zacharewski do not require general matters waivers because they do not have any personal or imputed financial interests in any pharmaceutical firms.

Because general topics impact so many

institutions, it is not prudent to recite all potential conflicts of interest as they apply to each member and consultant. FDA acknowledges that there may be potential conflicts of interest, but because of the general nature of the discussions before the committee these potential conflicts are mitigated.

With respect to FDA's invited guests, Drs. Krishna Ghosh and John Quackenbush report that they do not have a financial interest in, or professional relationship with any pharmaceutical company.

Dr. Kurt Jarnigan reports being employed full-time as Vice President, Biological Sciences and Chemical Genomics at Iconix Pharmaceuticals.

Dr. William Mattes reports being employed full-time by Pfizer, Inc.

William Pennie is employed full-time by Pfizer, Inc. and holds stock in Astra Zeneca and Pfizer.

Dr. Roger Ulrich reports full-time employment at Merck Research Laboratories and holding stock in Abbott Labs.

In the event that the discussions involve any other products or firms not already on the

agenda for which FDA participants have a financial interest, the participant's involvement and their exclusion will be noted for the record.

With respect to all other participants, we ask in the interest of fairness that they address any current or previous financial involvement with any firm whose product they may wish to comment upon. Thank you.

DR. KAROL: Thank you, Kimberly. Now Helen Winkle would like to welcome everyone.

Welcome

MS. WINKLE: Good morning, everyone. It is my pleasure this morning to be able to welcome each of you as a member of the Pharmaceutical Toxicology Subcommittee.

This subcommittee, which is a part of the Advisory Committee for Pharmaceutical Science, is important to the Center in addressing a number of questions and issues that come about due to the regulation of pharmaceuticals. This is one of five subcommittees of the advisory committee and really each one of these subcommittees has been very beneficial to us in helping to address various issues and concerns that we have, and helping us really develop the regulatory knowledge that is

necessary or the regulatory understanding that is necessary to maintain a strong scientific underpinning to our decision-making process. So, it is a really important group.

This is the first time the subcommittee has met. We look forward to a lot of interesting discussion over the years. Again, as I said, there is a lot that you all can contribute to us as we grapple with our decision-making processes. I appreciate all of your willingness to serve on this subcommittee and I especially appreciate Meryl for agreeing to chair this subcommittee for us. It is a big job and it will take time, and I appreciate her willingness to do that. I also want to thank all of the folks in the Center that helped make this subcommittee a reality. This includes Dr. Jacobson-Kram, Dr. Bob Osterberg and Dr. Sistare. So, again, welcome. We look forward to working with you. Thanks.

DR. KAROL: Thanks very much, Helen. Now the subcommittee is going to receive its charge and this will be delivered to us by David Jacobson-Kram.

Introduction to Meeting and Charge to Subcommittee

DR. JACOBSON-KRAM: Thank you.

[Slide]

I am relatively new to the FDA. I think this is my seventh week here, but this area is one of the things that drew me to the FDA. I think this is a very exciting time to be in toxicology and I believe with all my heart that this is going to be the future.

[Slide]

So, welcome to this meeting--the promise of toxicogenomics. What do we see as the future here? Using toxicogenomics, I believe we will be able to identify toxic responses based on mechanism of action. We will be able to identify those earlier in drug development. In the process of doing so, I think we will be able to use many fewer animals. By doing so, we will be able to optimize lead compounds early in development. We will have better extrapolation from animal data to human beings and ultimately, I believe, this will lead to faster development of safer drugs.

[Slide]

How about the challenge of toxicogenomics? Certainly the varied platforms and technologies--a lot of different companies are involved; there are different kinds of chips and these have to be

brought into some kind of uniform consistency.

Another big challenge is that correlations of expression changes and health effects are still evolving. We can document thousand and thousands of changes but we don't always know what they mean.

Finally, since everybody is coining new terms, I coined data "overlomics." This is one of the challenges with this field, the amount of data that it generates is overwhelming and trying to bring all that together and interpret it is certainly a challenge.

[Slide]

So, these are the questions for the committee, the charge: Should CDER be proactive in enabling the incorporation of toxicogenomics data into routine pharmacological and toxicological studies and in clarifying how the results should be submitted to the agency?

[Slide]

What should the present and future goals be for the use of the data by CDER, and what major obstacles are expected for incorporating these data into nonclinical regulatory studies?

[Slide]

Is it feasible, reasonable and necessary

for CDER to set a goal of developing an internal database to capture gene expression and associated phenotypic outcome data from nonclinical studies in order to enhance institutional knowledge and realize the data's full value?

[Slide]

Is it advisable for CDER to recommend that sponsors follow one common and transparent data processing protocol and statistical analysis method for each platform of gene expression data but not preclude sponsors from applying and sharing results from additional, individually favored methods?

[Slide]

What specific advice do you have for clarifying recommendations on data processing and analysis, as well as data submission content and format?

[Slide]

Today's program is divided into three topics. The first one is overview of toxicogenomics at the drug development and regulatory interface, and presentations will be by Drs. Woodcock, Ulrich and Pennie.

[Slide]

The second segment will be toxicogenomic

data quality and database issues, and the presentations will be by Drs. Jarnigan, Quackenbush and Ghosh.

[Slide]

The third part will be product review and linking toxicogenomics data with toxicology outcome, with presentations by Drs. Leighton, Levin, Mattes and Rosario.

[Slide]

Frank, I guess, will mediate the questions for the committee--

[Slide]

--and Dr. Karol will give us conclusions and summary remarks.

DR. KAROL: Thanks very much, David. Now I would like to have Janet Woodcock address us on the concept of no regulatory impact for nonclinical pharmacogenomics and toxicogenomics.

Topic #1 Overview of Toxicogenomics at the Drug Development and Regulatory Interface

Concept of "No Regulatory Impact" for Nonclinical Pharmacogenomics/Toxicogenomics

DR. WOODCOCK: Thank you and good morning.

[Slide]

What I would like to talk about this

morning is the whole issue of the emerging field of genetic information and also proteomic information and other allied types of information, and how that is going to play into the regulatory review process because the current regulatory review process that exists does not really formally recognize or incorporate this kind of information and, yet, it is coming; we are starting to see results in this area and so the question really does arise as to how do we, as a regulatory body, get this information; how do we deal with it; and also how we encourage the field to develop.

[Slide]

This is really about translation of innovative science to bedside medicine. This is about getting candidate drugs, lead compounds developed, get them through the process and to the bedside. How can we use new biological science that is emerging in speeding up this process?

[Slide]

Right now the new science of pharmacogenomics, and increasingly these other allied techniques, are applied extensively in drug development. They do have the potential--I agree with what was just said--to revolutionize the

process? Most of the data now is not seen by regulatory agencies, most of the data that are being generated, and partly that is out of concern for what we will do with it, to be very blunt. What interpretation will the regulatory agencies make of these findings?

Therefore, I think we need an approach that will enable free exchange of information, will help advance the science and technology along and will aid in the timely development of appropriate regulatory policies to apply to this kind of information. In the field of toxicogenomics we are seeking your help today in developing these policies.

[Slide]

Just for a brief background which I think you all know so I will go through this quite rapidly, but one of our problems as clinicians is the tremendous variability in human response to drugs. It is a huge barrier to using medicine effectively in human populations because you can't tell how people are going to respond.

[Slide]

There is variable effectiveness, and this isn't the toxicology side so much but it really

will also be related to animal models. So, for many drugs, if you leave aside antivirals and antibiotics and things that are directed at organism that aren't a human organism, the size of the treatment effect that we observe in randomized trials may be less than ten percent of the overall outcome measure, in other words, a very small amount of response. Many conclude therefore, correctly I think, that the effect of the drug is small, it is a very weak drug or the drug doesn't work.

[Slide]

If you look at it this way, if you look at a population basis, you see that you get a certain response in the placebo and if you use enough power in your study you can barely reach statistical significance often and show that the drug is more effective than placebo, but it is a very small difference.

[Slide]

If you define responders though--my slides in the book may not be exactly like on the screen, I am sorry--but if you find responders, then you can see that with the placebo you may get a little bit of response but for the drug you get a small

population that responds very well. We have seen this again and again in different areas. So, what we have here is variability. Some people respond to the drug and a lot of people don't respond to the drug. Our problem is that we don't know in advance who those people are so we have to expose a lot of people to get a small population responding.

[Slide]

In the same way, we get variability in the clinic in drug toxicity. If you look at drug versus placebo and you look in the PDR, or whatever, you see that every drug and even classes of drugs have a consistent pattern of side effects over the placebo. That is true for common events and it is true for rare events. Some of the wide effects can be attributed to the known pharmacologic effects of the drug and they tend to affect the population fairly uniformly, but may others are considered idiosyncratic. Again, the problem is we cannot predict which people are going to experience these side effects or experience them more severely. Therefore, currently in drug development as well as in medical practice we simply say oh well, this causes renal toxicity or liver toxicity and that is about as far as we get

and we watch for it. It is very observational and we really don't have a way often to say we should avoid exposing this group of people because they are more prone to this toxicity.

[Slide]

The good news is we think there is an inherited component, a genetic component to this variability in drug response. In other words, some of this would be predictable if we had more information.

I have two terms here, pharmacogenomics--there is quite a bit of dispute about what these terms mean so, please, this is simply for the purposes of this talk. I am considering pharmacogenomics to be application of genome-wide RNA or DNA analyses to study differences in drug actions. Pharmacogenetics, I am considering as looking at the genetic basis for inter-individual differences in pharmacokinetics and mainly that is driven by drug metabolism differences. But these two techniques can help us investigate this inherited or genetic component of drug variability.

[Slide]

In efficacy there are many ways to look at

this but there are at least three types of genetic variabilities that contribute to differences in effect of the drug, the beneficial effect. One is the diversity of disease pathogenesis. Of course, in animal models there are varying pathogenic pathways or actual diseases that lead to the same syndrome and often we don't have enough knowledge to separate those out and we expose everyone who exhibits a certain syndromic pattern. Some of them respond and many of them don't respond because they don't have the pathogenesis that would respond to that particular intervention. So, what disease?

Variable drug metabolism is a very important. What dose? People can have ten-fold differences in plasma levels based on metabolism. Right now we don't distinguish among those people. We give people a couple of ranges of doses and we hope they will all respond well.

Then, there are going to be genetically based pharmacodynamic effects. This has been studied, for example, in people with, say, differences in the beta adrenergic receptor. In people taking asthma drugs there may be genetically based differences in how well they can respond to a beta agonist. It has nothing to do with their

disease, but it has to do with other genetic variability underlying the genetic variability that they have, but still it may predict drug response. We are looking at that for some of the cholesterol lowering agents as well.

[Slide]

Drug toxicity, likewise there are genetic contributions to the variability in drug toxicity. One is that you may have a genetically based interacting state. You may have a long QT syndrome genetically, and you take a drug for some other condition that prolongs QT interval and you may be in trouble while the vast majority of the population has no effect from that. So, you have a predisposition to this toxic effect.

There may be differences in drug metabolism just like in efficacy. So, for toxicity there are some people, and we know this very well, who are actually overdosed significantly by standard doses of drugs based on their metabolic pathways that they have.

Finally, there are toxicodynamic interactions where you have a vulnerable subgroup. Again, it has nothing to do with their disease but they are simply vulnerable to some toxic effect,

some interaction. So, for toxicity, which is the main discussion at this meeting, at the level of the clinic there are genetic ways by which we could predict who is going to get a toxic effect.

[Slide]

But how important are these differences? That is sort of the skeptic's view. These differences exist. How much of human variability, for example, would be explained by genetic differences? Is this worth pursuing? Well, sometimes.

[Slide]

At the level of an individual a genetic difference in some cases can be determinative. I think this is the case both for toxic responses as well as for efficacy responses. More commonly at the level of an individual a genetic difference can highly influence drug response. It may make you much more likely to have a toxic response but not 100 percent, or it may make you much more likely to have or not have effectiveness in the drug metabolizing enzymes in your particular suite of drug metabolizing enzymes. You can really predict that you are getting the wrong dose or some individuals will get a toxic dose based on drug

metabolism. So, that can be very important.

[Slide]

But we have to recognize that many responses are going to be an emergent property of multiple gene products that are interacting both with each other and with the environment, environmental factors. So, that is where we may have to look at patterns. That is where proteomics and other things come in because this will be more of a systems issue than a single factor that is determinative or highly predicted.

[Slide]

I like this pyramid, which is from Science recently, which talks about the different levels if we are looking at these things. At the very top is the organism, the mouse or the rat or the monkey or the human, and we are an interacting system of many, many subsystems. When you are looking at genetics you are down at the bottom; you are only looking at a piece and it contributes up; the same with proteomics and many of the other studies. This is where the data that David was talking about comes in because we have to take many snapshots of the organism at many different levels to understand what is really going on.

[Slide]

Currently drug development is satisfactory but it is very expensive and we find out things very late in drug development that would be much better to find out early. We are able to determine whether drugs are effective or not. I can tell you that the Center for Drugs does not approve drugs that are not effective anymore--

[Laughter]

--but we use a population basis. So, what the public asks us today is more is this going to work for me? They don't really care if a drug works hypothetically in a population; they want to know is this drug going to be effective for me. We can't tell people that right now when we approve a drug.

The same with drug toxicity. As you all know very well; you are more expert in this than I, the determination is observational. It is based on exposing animals and the human is very similar. We expose the human but we just don't go up to the toxic doses we do in animals, and we see what happens. Again, when we put that drug on the market and it is being sold we can't tell a patient, individual patient, you are the one; you

are going to get the catastrophic side effect; you are going to get this bad side effect; or, you are going to do just fine on this drug. We do not have that kind of information. Whatever guiding information we give to clinicians is very crude--avoid in renal failure or something like that; it is a very, very crude level. Right now carcinogenic and reproductive toxicity potential of the drug is based on the in vitro and animal studies and, again, we do pretty well on this but we can't tell people for sure.

[Slide]

What potential uses do we have for this genetic information in drug development? Well, David has already talked about this a little so I will go through this quickly. Obviously, improving candidate drug selection is very important given the cost of drug development. Developing new sets of biomarkers for toxic responses, first in animals and then in humans, eventually with the goal of minimizing animal studies and, yet, having better predictability from our preclinical work. At the clinical level, predicting who will respond and who will have a serious side effect--this would be wonderful. Also to rationalize drug dosing based

on the genetic substrate of the individual.

[Slide]

In sum, we can all, the biomedical community in general can pull this off. We can expect for the next decade or two to move from the current empirical process--which is what drug development right now really is; it is not a mechanistic, predictive type of process--to a mechanism-based, hypothesis-driven process for the triumph of rational science in biology, which is something we haven't really been able to achieve yet. This would result in a lower cost and faster process that could result in more effective and less toxic drugs, albeit they would be indicated for smaller groups of people because we would know from people's genetic and other information who was going to respond.

So, the potential of this is tremendous. I agree with David, I have no doubt this is going to happen. It is just how soon and how many bumps we are going to encounter in the road. Frankly, today one of the things you are going to discuss is one of those bumps and how do we deal with one of those obstacles effectively.

[Slide]

So, that is the question, how can this new technology be smoothly integrated into the drug regulatory process? How can we do that?

[Slide]

Right now our legal requirements, which are driven by the Food, Drug and Cosmetic Act, require that we evaluate all methods reasonably applicable--this is in the new drug application--to show whether or not such drug is safe for use under the conditions in the proposed labeling. So, all methods reasonably applicable about safety. For effectiveness, that we look at adequate and well-controlled trials to show that the drug will have the effect it purports to have under the conditions of use.

[Slide]

For the investigational new drug application, the IND, there are submission requirements in our regulations. They state that you have to submit the pharmacology and toxicology information on the basis of which the sponsor has concluded that it is reasonably safe to conduct the proposed clinical investigations. That is what the regs say.

[Slide]

About the NDA submission the regs say that for nonclinical studies you must submit studies that are pertinent to possible adverse effects. Obviously, when these regs were written we did not know about this kind of information that we are talking about today.

For the clinical you have to submit data or information relevant to an evaluation of the safety and effectiveness of the drug product. So, relevant.

[Slide]

The issues that need to be resolved are when and how to use developing pharmacogenetic information and related information in regulatory decisions. When is the information reasonably applicable, pertinent or relevant to safety? That is really one of the questions. And, under what circumstances then is submission of this information about a candidate drug to FDA needed or required? Under what circumstances?

[Slide]

We have already developed somewhat of a plan on this but what we are here today for you help fill in some of the details I think. We discussed this plan or proposal with the FDA

Science Board and received some endorsement, but the proposal was at a very high level without detail filled in.

What we propose to do is we will establish policies on pharmacogenetic data and we will have a policy on what type of data is required or not required to be submitted; what type of data are appropriate or not appropriate for regulatory decision-making. This is the kind of information the sponsors need to have.

[Slide]

What about submission requirements? I have to stress we do not have a policy right now. We are working on one and we will go through a public process, as I will describe, but we would decide whether or not submission of data were required based on interpretation of the regs and the statute that I quoted above. It is clear right now, that without any interpretation, that any data actually used in protocol decision-making in people needs to be submitted. That is probably true with animals too. If you are going to select animals on genetic data, and so on, and manipulate them in some way in the protocol, or whatever, that would be obviously required.

In addition, it is clear and may have happened, I am not sure, that sponsors may submit data to FDA to bolster a claim or their scientific position about something. For example, people may want to explain why a finding in a certain animal species is not relevant to humans and they may wish to submit a variety of genetic data to show that the relevant genotype, or whatever, is only within that one species, or whatever. But for most results, as I have here, submission not required. This line is the line that we have to work on and FDA is working on that.

[Slide]

The thing about submission of data, if submission is not required, how is FDA going to develop a knowledge base about the field? This is the conundrum we are in. So, we will be requesting voluntary submission of results, and this is where "no regulatory impact" comes in. Results would not be used in regulatory decision-making. We really do need to hear about emerging results as this information begins to be used routinely.

[Slide]

But how would we give this assurance? When would FDA use the data for regulatory

decision-making? I have to stress that this is sort of a working proposal that we are thinking about. FDA will apply a threshold determination to the data that is submitted. Okay? Data that is submitted voluntarily would already be in the category of "we would not use that for regulatory decision-making." All right? Data submitted by a sponsor to make a case, obviously we would use that in regulatory decision-making; the sponsor would be requesting us to use that in regulatory decision-making. So, there are really three categories of data that we are talking about here.

[Slide]

What we are proposing, and this is just a work in progress, is that the information would have to have risen to the status of being a valid biomarker. In other words, when the meaning of the genetic test is well understood and of known predictive value, then results from testing animals or patients should be submitted to FDA. In other words, it would be required. That would be the required submission threshold. This clearly could be whether we use this for a regulatory decision-making threshold because we don't use information for regulatory decision-making if it

doesn't really have meaning yet.

The problem with a lot of the genetic information, as you all know, is it is currently being generated and we don't know what it means. In a sense, we know what it means in a genetic sense but we don't know what it means in a predictive sense. We don't know what it will imply and, therefore, we shouldn't be drawing conclusions about it. Research or exploratory tests, in fact, are not suitable for making decisions on safety or efficacy of a drug. They are not yet suitable.

[Slide]

What we are planning to do is develop this threshold and these policies using a public and transparent process with advisory committee oversight. While I know today the main focus of the effort is to talk about the standardization, and so forth and so on, this discussion toady before this advisory committee is what will help feed into the policies as we develop them.

[Slide]

What we plan to do is publish a guidance for industry that would have a decision tree for the submission, what is required to be submitted, and also a decision tree for whether things would

have regulatory impact or not, whether the data would have regulatory impact. Is everybody following me on that? Is that clear?

What we do when we do a guidance is we will publish a draft. We hope to publish that in August. Then we will have extensive public comment on the draft and we will probably have a workshop after that draft is published so that people can react and we can have extensive input. Then we will probably have more advisory committee discussions about the draft. We will also establish an interdisciplinary pharmacogenetics review group that would provide a centralized review of this information. We have a carcinogenicity committee that looks at all the carcinogenicity studies to provide consistency across the Center. We will do the same thing for this type of information so we will have a centralized review and this body could also work on ongoing regulatory policy development.

[Slide]

As part of today's discussion, we will be working with the advisory committee and talking about our work in the private sector on the standardization issues. Obviously we will never be

able to use this information in regulatory decision-making if it isn't standardized in some way so we can understand what it means, one platform to another. Standardization is really one of the basic efforts you have to go about working on when you work on various biomarkers so that you know what the results in one lab mean compared to another lab. As I said, we will also issue a guidance, a separate guidance on the format of the submission and the data, in other words, how we would like to see the data, and that is going to be discussed today.

[Slide]

What are some examples? These might be controversial so let me say this is just the working proposal and we may modify this even in the guidance. What about genetic information generated in animals, in toxicology studies? We don't know what would be required to be submitted right now to the FDA because we don't know of anything that we would understand well enough that it would be considered valid by a marker to be submitted. All right? That is going to change over time, we all hope, but that is the state we are seeing right now.

We are definitely interested in voluntary submissions and we are not seeing very many. Again, as I said, to explain an animal toxicity finding, that is really up tot he sponsor, to submit that and I think people have submitted things like that.

[Slide]

We have been asked this question in toxicology for animals, for cells, for people, what if you are doing a screening study, an expression study and you are looking across a genome and what if you expose this cell, animal or person to drug and you see increased expression of an oncogene after drug exposure, or maybe many oncogenes?

Well, we have looked into that, and I hope Frank talks about that a little bit or someone talks about that, but we looked into that because we were explicitly asked and this is the kind of thing people are worried about. What we find is that in some studies that have been done many common drugs that are given at high dose can elicit this finding in toxicity studies. Of course, these proto-oncogenes weren't really put in the body to cause cancer. They are used in development or repair and other types of physiologic actions and,

naturally, they are going to be turned on after injury, during development and so on. So, this encapsulates I think what the sponsors are worried about, that they would find something like this. They would submit to the FDA and their drug would never see the light of day basically. But this shows, I think, the value of looking across a broad range of studies, understanding what is going on and having a scientific database because we are able to put these fears at rest very easily simply by looking at what has been done. But this question will come up again and again as we start really probing and finding out what is turned on when animals or cells are exposed to drugs.

[Slide]

I just put this in although this is clinical pharmacology. People may want to genotype or phenotype trial subjects for their isoenzyme polymorphism for drug metabolism. Now, in this case, the value and meaning for many of the isoenzymes is very well known and it is relevant to assessing outliers in pharmacokinetic studies. It is relevant to looking at the people who experience drug toxicity and see if they were effectively overdosed in the study due to their genetics. So,

this kind of information should be submitted to FDA, should be evaluated by us. In fact, recently it was put in a drug label for a drug, and should probably go in more drug labels. I don't think there is a lot of fear about this in the industry or anywhere because we all know what this means and the value of this information.

[Slide]

This, again, is a working proposal. What if you gather a bunch of screening genomic data in patients during a clinical trial, does that have to be submitted to the FDA? Our current proposal would say no. But what if you analyzed the data and you saw a potential correlation with an adverse event? What would FDA do? There have been very exaggerated fears out there that we would say, well, you can't give this drug to people who might have this genotype, and so forth. How would we interpret this?

Well, it is basically simply a potential biomarker, and the way we look at those is that you need a lot of evaluation in additional trials and diverse populations because I think one of the things that is going to happen in humans, other than animals, is humans are a very outbred

population obviously and there is going to be extensive variability in the findings. We have already seen this in humans. You are laughing but we are--we are becoming more outbred every day. There is extensive variability in the frequency of certain genotypes and, therefore, the clinical impact of these findings depends on what human population you study. So, simply because you find it once in humans doesn't really mean a whole lot except that it might be of interest.

[Slide]

In summary, I think that pharmacogenomics really does hold great promise for drug development and for rational therapeutics, which is really the goal in the clinic, to really understand who we are giving the drug to and be able to predict what the effect will be. In fact, use of this technique is increasing. It is actually very widespread in industry right now. What we need is free and open exchange of results between the industry and the FDA to ensure the appropriate development of regulatory policies.

[Slide]

Concerns about how the data will be used by the regulators has stifled this exchange to date

and is continuing to. FDA will develop clear policies on the use of pharmacogenomic data in regulatory decision-making both for toxicology and clinical. And, I think we all look forward to the advances in medicine and health that these techniques, I believe, are sure to bring eventually.

I thank the committee for its work. You will be making some steps today towards making this come about. Thank you very much.

DR. KAROL: Thank you very much, Dr. Woodcock. Are you available for questions from the committee? Would any of the committee like to ask a question?

[No response]

Thanks very much. We will move on then to our next speaker and, unfortunately, Dr. Ulrich isn't with us today because of the death of his father. So, we will have the following speaker now, and that is Dr. Pennie who will talk to us on a perspective on the utility and value of expression profiling data.

A Perspective on the Utility and Value of Expression Profiling Data at the Drug Development Regulatory Interface and ILSI Experiences with

Cross-Platform Comparisons

DR. PENNIE: Thank you very much.

[Slide]

It is my pleasure to speak to the committee this morning, and my privilege to represent a working committee organized under the auspices of the ILSI Organization, which is a consortium effort amongst industrial organizations, academia and government to address some of the technical challenges and share some of the learning on these emerging technologies related to genomics applications and risk assessment.

[Slide]

This committee has been in existence since mid-1999. When the committee was formed, what I have here is a slide of some of the challenges the membership believed were facing the advancement of these sciences, the first one being a lack of publicly available databases to help put experimental data in context; the second one being a lack of validation of the available technologies; a lack of comparable tools, methodologies and study designs; a lack of robust and consistent tools for data analysis; a lack of fundamental knowledge of how gene products relate directly to toxicity and,

in particular, the relevance of single gene changes. When I speak of genes in the context of this presentation, I am talking largely about genomic changes where we are measuring basically the induction of gene expression or repression as a consequence of a compound treatment. So, we are not dealing in this committee's work at this stage with a variable response which may be a result of genetic variability. Certainly, the last comment here, uncertainty about the regulatory environment, was a comment which I think was raised quite eloquently in Dr. Woodcock's presentation, and certainly having a committee like this before us today is an opportunity to broaden the dialogue in this area.

[Slide]

So, for those of you who aren't familiar with it, the ILSI Health and Environmental Sciences Institute is a non-profit research and educational organization which provides an international forum for scientific activities. These are largely experimental program-based activities. The ILSI organization enjoys participation from industry, primarily the drug industry, the agrochemical and chemical industries and also from government and

academic researchers and advisors. The organization runs research programs, workshops, seeds databases, forms expert panels and actively pursues the communication of its findings through a publication strategy, and has a reputation for focus and objectivity.

The ILSI organization is not a trade body. It has specifically in its charter that it does not attempt to directly influence the setting of regulatory positions or policies. Instead, they try and provide a basic and fundamental understanding of evolving technologies for how these technologies may be used.

[Slide]

As I said, the committee was formed in 1999. As it stands, it has a membership of around 30 companies, an international-based membership, including government participation from labs such as NIEHS, NCI, NIH, NCTR and others. We also enjoy a very active participation of a group of academic advisors who sit on the steering committee of the organization.

[Slide]

Our objectives were to evaluate experimental methodologies for measuring

alterations in gene expression, alterations as a consequence of compound treatment. Other objectives included the development of publicly available data to allow the beginning of discussions on relevance of findings and issues around the development of databases.

Particularly, we charged ourselves to contribute to the development of a public international database linking gene expression data and key biological parameters with the goal of determining if known mechanisms and pathways of toxicity can be associated with characteristic gene expression profiles or fingerprints, as they have come to be known in this field, and if the information can be used as the basis for mechanism-based risk assessment. So, we are talking primarily about an application in a preclinical setting here.

[Slide]

Here is a time-line of where the committee has come from and where we are at the moment. In early 2000 the committee initiated an experiment program which focused on three areas of toxicology for further evaluation, those being hepatotoxicity, nephrotoxicity and genotoxicity. We also formed a

database working group to look at issues around data capture, storage and transmission. We initiated a collaboration on database issues with the European Bioinformatics Institute early in 2002. You are going to hear a little bit more about that initiative at the end of my talk and in Dr. William Mattes' talk this afternoon.

Just last week, in fact, we held our first public meeting on the application of genomics and risk assessment, in the Washington area, and invited a large number of scientists from the regulatory and academic communities to join with us in discussing the progress of the committee to date and future opportunities for sharing of learning as we move forward with these initiatives. We also have an aggressive peer-reviewed publication strategy which will take us through 2003 and the early part of 2005.

[Slide]

Let me tell you a little bit about what the actual deliverables of this committee are. The program mechanism was, as I said, to organize ourselves into a series of working groups to focus either on experimental research in the areas of hepatotoxicity, nephrotoxicity and genotoxicity or,

as I articulated, to begin discussions and planning around contributing to an international database on gene expression changes.

[Slide]

Our experimental design feature basically profiling well-studied compounds in the literature with known toxicity profiles and biological parameters. We investigated temporal relationships and the effect of dose on gene expression changes and an opportunity afforded by the committee, as you will see, is that given the broad membership and broad access to numerous technical platforms, we have the opportunity to look at some technical details of the technology, including variability and operating procedures that may vary from one laboratory to another.

[Slide]

I have made a list of the objectives we set up at the beginning of the committee's activities to try to give you an understanding of what our status is. For the first objective, to evaluate methodologies, we have developed protocols within our member labs and within the committee as a whole to evaluate profiles of specific prototypic toxicants. We went through an exercise of

distributing RNA samples to public and industry labs for microarray-based gene expression analysis. This allows us to consider variability that may take place both in in-life studies and inter-lab variability when different labs are profiling the same material. We evaluated the influence of specific experimental conditions on data variability. These may be technical experimental conditions such as the way that the apparatus is set up for the experiment. Those issues are still being looked at. We have utilized the outcome of experiments and data analysis to stimulate discussion of what the best practices may be for these applications.

[Slide]

A second objective, to contribute to the development of international databases linking gene expression data and key biological parameters, will be discussed in a little bit more detail briefly at the end of my talk but also in Dr. Mattes' talk, but effectively, we have been in discussion with a large number of stakeholders on data formats for microarray storage and transmission; building database structure to include the incorporation of standard toxicology endpoints in preclinical

studies; and a drive to make these databases and the data within them available in the public domain actually before 2004 but, we expect, in the course of this year.

[Slide]

A third objective, this is where we start to focus on risk assessment, is to determine if known mechanisms and pathways of toxicity can be associated with characteristic gene expression profiles and if this information can be used for risk assessment.

So, as I have said, we have developed gene expression datasets on well characterized toxicants and are at various stages of data mining and data evaluation to characterize the mechanistic information that can be gleaned from such studies.

[Slide]

I will very briefly give you an outline of the three working groups, then I will try and give you, for each one of them, some of the interim conclusions the working groups have reached with regard to the technology and its applications.

Our nephrotoxicity working group worked on three prototypic nephrotoxicant compounds and had in-life studies conducted at a single site to

prepare material in vivo for the analysis of these compounds' effects on transcription profiles in lab animals. In this case it was in rats. There were eight participating labs who were involved in taking the material from the in-life study, preparing and analyzing it using gene expression analysis technologies. These technologies, including multiple technical platforms, the microarrays produced by organizations such as Affymetrix, Incyte, ClonTech and Phase-1 and also the use of custom cDNA microarray platforms which have either been generated in academia or in the labs of the participant organization, and pooling all this together gave the opportunity to compare inter- and intra-lab variability, cross-platform variability and the ability to replicate the in-life study.

[Slide]

So, the interim findings were really an ability to recapitulate the data on standard tox endpoints for these compounds. In other words, we were able to replicate what was known about the more traditional tox endpoints in the rat species for these compounds. Transcriptional analysis yielded strong topographic specificity and some

mechanistic information about the mode of action of the compounds.

Where we had individual gene expression changes that were of interest to the committee, we did confirmatory analysis using alternative methodologies. All of these were positive and will be extended to investigate potential biomarkers of nephrotoxicity in preclinical species.

The frequency of individual animal transcript changes was reduced in non-responders and increased in cases of severe toxicity. In other words, there was a direct linkage between the magnitude of gene expression changes and the onset of toxicity.

We, not surprisingly, found that the use of pooled RNA samples may have a dilutional or skewing effect on the interpretation of genetic response, but at the stage these programs were initiated cost was a major factor in being able to take these programs forward and pooled samples were analyzed in the initial stages.

The group has concluded that these technologies have at least equal sensitivity to traditional toxicology endpoints in terms of detection and an enhanced opportunity to resolve

some mechanistic information.

[Slide]

I will move a little bit more quickly through our second working group. You have the tenor of how the groups are organized. The hepatotox group worked on two test compounds but they performed independent in-life studies to look at the effect of different sources of in-life material and in-life studies on data analysis. They had 14 participating laboratories in the analysis of the material, again performing analysis on multiple technical platforms. The use of 14 industrial labs on two test compounds and two in-life studies gave a truly unprecedented opportunity to look at issues related to variability.

[Slide]

Their findings were, again, the expected outcome with regard to the in-life study replicating what was known in the literature about these two compounds. Within a given technical platform, in other words, using a single microarray platform such as Affymetrix, there was a high degree of concordance, greater than 90 percent, in the direction of the of the gene expression changes

across samples analyzed in different labs, but lesser concordance was observed when identifying probes or individual genes that were regulated above or below a certain threshold for all datasets, for example, a cut-off of greater than 4-fold to regulation. This result may be attributable to differences in data capture algorithms or data analysis methodologies across labs.

Dose-related response was observed in these experiments, and for one of the compounds under study, methapyrilene, agreement was found across all platforms with good but varying degrees of congruence in the results.

Now, the field of data analysis for gene expression changes is very much on a logarithmic scale in terms of its advancement and since this slide was made there have been some strides forward in this particular working group in reconsidering their methodology for data analysis and, in fact, we believe that if you limit your data analysis to genes that have a very high degree of statistical rigor around the expression change within an individual lab, then the cross-lab variability is significantly reduced.

[Slide]

A slightly different approach was taken by our genotox working group which conducted their assessments in cell lines, the mouse lymphoma p53 null cell line and the human TK6 cell line which is p 53 competent. They run their gene expression profiling experiments in concert with standard genotox testing regimes to look for direct-acting mutagens and clastogens microarray analysis on the material prepared from the cell lines and, again, multiple platforms were used for the comparisons.

[Slide]

Their conclusions were that gene expression changes less than 3-fold were very common in all studies even at highly genotoxic concentrations. So, concerns around the over-sensitivity of the technology appear to be unfounded, at least with the limited dataset generated by this group.

Array technology in fact may not be as sensitive an endpoint as the more standard genotox testing battery which is currently in use in the industries, but gene expression changes have the advantage of possibly allowing us to distinguish mechanistic classes of genotoxic compounds. The

strong push from this group is that standardization of analysis and control of experimental variables, as we have discussed already this morning, pose challenges to data comparison and interpretation.

[Slide]

the committee-wide data findings, to summarize, are that application of microarray technology has all the usual sources of experimental variability you would encounter in a biological experiment, with the additional complexity, which can come from a number of areas, such as differences in the protocol for the harvesting of the mRNA sample; differences in protocols or conditions for the hybridization of the RNA sample to the microarray platform; importantly, differences in the way the genes are recorded by manufacturers on their individual technical platforms. In other words, gene X may not equal gene X between two different technical platforms--different specific nucleotide sequences within probe sets across different technical platforms. In other words, even if gene X on platform 1 does equal gene X on platform 2, the precise sequence used to make the detection may be different and be subject to different hybridization

kinetics, for example.

Clearly, a big issue is that all these are not made equal and there is not a direct correlation for the gene sets on one manufacturer's array to the gene sets on another's. It is important to monitor the effect of signal to noise ratios; analysis setting on the machinery used to make the detection; keep a hold of false-positive and false-negative rates statistically to make sure you are not putting too much weight on background noise in an experiment. Clearly, there are a large number of different analytical tools that take the raw data from these experimental platforms and convert them into a subset of gene changes for further investigation. There are significant differences in the methodology for getting at that analyzed short list that can have a fairly significant effect on the interpretation of a given experiment.

[Slide]

This slide I think just summarizes the opportunity that was afforded to the ILSI membership and, by its charter, is afforded to anyone in the public community or regulatory community who would like access or discussion on

the data. This slide basically then captures where we have had an opportunity to look at variability issues, be it the in-life variability, variability in in vitro experiments, intra-lab platform replicate variability, and so on and so forth.

[Slide]

Very briefly then, we heard this morning about a data overload in genomics technologies. What was once promised us as a great advantage and a step forward for these technologies and the rapid accumulation of very high density of information turned pretty quickly into one of the biggest challenges for people who dealt with the data in terms of managing, storing and interpreting the many, many millions of data points that can be generated from even a single experiment.

So, in recognition of this, the ILSI committee, as I said earlier, engaged in a collaborative effort with the European Bioinformatics Institute on building and enhancing their existing ArrayExpess database platform, which houses array data from multiple technical platforms, is compliant with the internationally regulated standard for the minimal information required for a microarray experiment and,

importantly, has been extended to incorporation of toxicology endpoint data into a microarray submission. In fact, there has been the evolution of a new microarray data standard, called MIAME-Tox, which is the subject of one of this afternoon's presentations. As I said earlier, the database is largely functional. The tox component of the database is expected to be rolled out to the public domain sometime in the course of 2003 or early 2004.

[Slide]

The complexity of such a database is hard to get across to people when you are trying to capture not only the data itself but the experimental conditions that were used when the experiment was performed, and also additional biological information that is important to put the transcriptional data in context. So, we have within this database schema the opportunity to store information on the sample pool, the way the material was extracted and prepared, all the experimental conditions around the generation of the gene expression data and link that directly to various biological endpoints, such as traditional pathology, biochemistry or clinical chemistry

endpoints.

[Slide]

Winding down this presentation, the program status for 2003 for the ILSI committee is that we have completed the data analysis, effectively completed the data analysis from current studies. These were what we considered the Phase 1 studies that we initiated in 2000. We have completed an interim review and, in fact, published an interim conclusions document which is available from the ILSI web site.

We had, as I said, an invitational worship just this last week to discuss the interpretation of the committee and take forward issues around the application of genomic data in risk assessment. We valued very much the dialogue between the committee, the academic sector and various invited participants from FDA and other regulatory agencies and, indeed, at that meeting recognized the importance of moving forward in the ILSI committee of having some steerage from the FDA as to what were important questions for us to answer. So, as a result of discussions last week we invited Dr. John Leighton to join the steering group of that committee and he graciously accepted.

Our collaborations are to continue to analyze issues of variability. We have internal efforts within and across participant labs to look at variability of analysis, and we are also grateful for collaborations we have initiated with external organizations, such as Affymetrix and Rosetta Informatics, to help with consensus on the important issues around the methodology for analyzing data.

As I just showed you, the EBI database continues to be supported by the ILSI committee and the evolution of standards from microarray expression data exchange is high on our radar for important activities moving forward.

[Slide]

White papers on interim findings, as I said, are available right now on the ILSI organization's web site. A series of peer-reviewed publications, including back-to-back publications scheduled for the fall, initiated in spring 2003 and take place through 2004. We are in the process of writing up the minutes from our invitational workshop; continue to move forward with EBI and ongoing discussions, such as the one we are having this morning and this afternoon, on the application

of these methodologies to risk assessment and the best practices that need to be put in place for best interpretation of the data.

[Slide]

Here is my final slide. I have tried to list here what I think are the opportunities that are afforded to all interested parties, and particularly this committee on the application of genomics to mechanism-based risk assessment. I think this particular committee has an unprecedented opportunity to compare multiple platforms analysis methodologies and inter-lab variability issues. Remember, we were able on this committee to harness the infrastructure of 30 or so large pharmaceutical and other industry companies, comparing results across multiple technical platforms that no one individual organization would have been able to do by themselves.

That has also given us the opportunity to sit down with colleagues across the industry, academia and the regulatory agencies to discuss where we are going with improving methodologies. We have the opportunity to engage database experts and to seed a publicly accessible and linkable database, and to ensure that such a database is

able to incorporate or link to toxicology information.

What I didn't say earlier is that a key issue was that that data would be transportable to other databases that may evolve in the academic or public sector and, as such, could be very much a partnering opportunity as the data begins to evolve in pockets amongst the emerging databases.

It has given us the opportunity to contribute to discussions such as these on the appropriate application of the technology and, importantly, these discussions can be based on shared experience rather than perception around what the technology may or may not do. I think it is important to promote appropriate usage in an industrial setting to maximize the usage of these approaches in a holistic safety assessment process.

Dr. Woodcock said this morning that there are a number of fear factors which we have to overcome to get the best usage of this technology. Some of the biggest of those to overcome are actually those that exist within the industries themselves. Not so much fear of how regulators are going to analyze the data, but really just fear of doing the experiment in the first place. It is a

fairly standard approach in toxicology and certainly in risk assessment experiments that you should not conduct an experiment if you are not confident you are going to be able to interpret the data. You have to think harder about experimental design if you find yourself in that situation. So, clearly with emerging technologies such as these, there is a fear within the industries that we are going to generate data that we are not fully able to understand and, therefore, a rather conservative approach can be adopted to not do the experiment and not advance the science. So, hopefully, today's discussion is part of the process of trying to instill courage, both in the regulators and the regulated, to move these very promising technologies forward.

So, with that, I am happy to take any questions if there are any and, again, thank the committee for the opportunity to come and participate in the discussions today. Thank you very much.

DR. KAROL: Thank you very much. Are there questions from the committee? Yes?

DR. BROOKS: Talking about the interactions between your working groups, you had

stated that at least on some level there was concordance across platforms since you are using multiple platforms. Any numbers or percentages with respect to those platforms within the working groups?

DR. PENNIE: It is very dependent upon how you do the analysis. For example, some of the early figures which we reported at the Society of Toxicology meeting two meetings ago were based on a less than critical assessment of the statistical rigor of an experiment within an individual lab, if you see what I mean. So, those were very disappointing figures I think, that even what we thought was a well controlled experiment may give you, you know, less than 20 percent agreement in the gene list for an individual experiment. But, rather than give you a number right now, I would say watch this space because we have some very encouraging results, particularly from the hepatotox group where a more rigorous analysis gives a much more comforting result even with the number of gene expression changes that stand up to that rigorous analysis give you a much shorter gene list at the end.

DR. BROOKS: So, higher statistical rigor,

you think, will give you higher concordance across platforms?

DR. PENNIE: I think it may, but also a greater understanding of exactly what the annotation issues across platforms are, which is part of that rigor exercise. There is no point in trying to compare gene X to gene X on another platform if, in fact, they are not gene X.

DR. BROOKS: One other quick question, what do you think the relative contribution of each of the additional variables associated with microarray data is that you had listed on that one slide, in the hopes that some of them may actually not be as significant and some will be more significant, so we know where to focus our efforts?

DR. PENNIE: That is a good question. I think one in particular for the Affymetrix platform is the PMT setting on the detection apparatus. What I think that is likely to skew the results for is really borderline calls between present and absent on a given microarray. In other words, you will have a different size of gene expression shopping list from one experiment to another but it will be overlapping, and there is an area of sort of noise versus signal that may be lost in an

inappropriately calibrated machine.

DR. BROOKS: From this data, do you think you can do some kind of a transformation analysis to assess the contribution of those sources?

DR. PENNIE: That is possible. In fact, those and other issues were part of the collaboration we engaged in with Affymetrix directly to try and identify some of those sources of variability.

DR. KAROL: Some of the anticipated benefits from this technology is increased sensitivity and mechanistic insight. Can you comment on your findings relative to that? DR. PENNIE: Mechanistic insight I think is something that practitioners of this technology in an industrial setting have been very confident about if you run a well-designed experiment that is not just generating a shopping list of gene expression changes. In other words, if you believe that you have a hypothesis to prove that a particular toxicant may be operating through a particular pathway, then you can remove some of the experimental variability by using small molecule inhibitors or transgenic models, for example. Those are extraordinarily powerful combinations of

multiple technologies and have some very compelling examples of an increase of the mechanistic understanding of a compound's action. So, I am not pouring a lot of comfort in the committee that in a risk assessment sense these technologies will be adding value.

DR. KAROL: Did you gain any mechanistic insight from your studies?

DR. PENNIE: Indeed, we did. Actually, there are a couple of manuscripts in preparation and, in fact, we came up with some new mechanistic insight on the particular toxicants we have had under study that will be published in the peer-reviewed literature.

DR. GOODMAN: Before getting too much into the question of effect of experimental treatment, could you address the issue of variability in controls? How consistent are the controls, and are there differences in terms of variability depending on which platform is used?

DR. PENNIE: Yes, that is a good question. So, if you compare control data with an individual set of protocols performed within an individual lab the results are reasonably consistent, stand up to what you would expect from that kind of an

approach. The challenge is in comparing control data from one lab to another. In fact, until we get a better handle on experimental methodologies and sources of variability, particularly in the analysis, it is not too surprising to practitioners that control data from different sources actually gives a greater amount of difference than control and treated within an individual lab. So, that is a significant source of variability. But within an individual lab control data tend to be pretty tight.

DR. HARDISTY: When you selected your compounds for this test for nephrotoxins or hepatotoxins, did you have any that were not known to be nephrotoxic or hepatotoxic to look for false positives?

DR. PENNIE: Yes, that is a good question. Instead of doing it that way, what we did, particularly in the nephrotox study, was that we harvested other tissues, other than kidney, so that we would be able to look. In other words, the nephrotox non-kidney tissues were used as negative controls for the hepatotox experiment, if you follow me. It wasn't a rational part of an individual working group design but that material

is made available for the other groups to look at different tissues than the classical site of action.

DR. WATERS: On the slide at the top of page seven you use the term topographic specificity, which I think I like very much. I would like for you to just expound on that thinking.

DR. PENNIE: Okay, that one is referring to the nephrotox working group. We were specifically using compounds that are at a different site of action in the kidney. After the microarray expression experiment had been performed we were able to use other technologies, such as in situ hybridization to show that the changes in expression were actually associated with the site of toxicity.

DR. ZACHAREWSKI: At the meeting last week there was an interesting discussion regarding liability and culpability in terms of the historical aspects of data reanalysis years after the fact to identify that. I was wondering if there was an opportunity--I will take the opportunity to ask whether you have any comments and see if there is any clarification for FDA

because I don't know if there was an opportunity for FDA to respond to that as well.

DR. PENNIE: That is a very good question, Tim. I appreciate it. I think there are two challenges here. One is that as the field evolves we will collect more and more data on the relevance of individual transcriptional changes and have more and more mechanistic understanding of various tox endpoints. So, there continues to be an onus on the organization that has generated the data to reflect back on their findings in the light of advancements in research to make sure they did not observe a toxicological flag that has been subsequently validated. So, that is one challenge and I don't know if we will get some response from our FDA colleagues or not this morning.

An even bigger one for me though is we will just spend some time discussing how variations in your analysis methodology can give you a different result. So, clearly, you can analyze an experiment and think you have the answer, and not only can the science move on but the analytical approaches can move on. So, somewhere along the line you have a lot of opportunities to not be picking up on what could be a potentially

significant finding. So, for me, this all boils down to a comfort around individual genes as not being an appropriate level of scrutiny for taking these technologies out of context in a risk assessment paradigm. If we can cross that bridge and understand that we have to have a lot more meat and bones to a risk assessment argument than single gene expression changes, I would hope that we would find ourselves in a very sensible place with regard to those issues. But, certainly, comment from our FDA colleagues would be extraordinarily valuable.

DR. WOODCOCK: Could you explain the question a little more clearly because I wasn't at the prior meeting?

DR. ZACHAREWSKI: Well, the discussion centered around the fact that, you know, if company A generated microarray data and they analyzed it to the best of their extent at that point in time and that data was then deposited within a database, ten years down the road if somebody else reanalyzed that data with the new technologies and the new information there was discovery associated with an adverse health effect, would the company now be liable as a result of that and, I guess even greater than that, be culpable associated with

that?

DR. WOODCOCK: Right, Well, I think there are two separate trains of thought here. One is sort of the regulatory train and then the other is product liability, which is a much less predictable and maybe science-driven process. In general, I would say though if you look at drug development, you are looking as positive control things we know, known toxicants or whatever. We, in the course of drug development--we, meaning the community involved in drug development, find these things because we expose animals. We are going to continue to do, in other words the routine studies both in animals and in humans, and we will find most of these. I think the ability to predict rare, catastrophic adverse events in people is going to be one of the last things to happen. The other kind of events we are going to find out during drug development so it wouldn't be like you would be clueless and you would have a drug on the market and you wouldn't know, I don't think. So, from a liability standpoint, you have already gone through the vulnerable period, which is when you are in drug development and you don't really know and you are exposing humans for the first time.

But, of course, in the courts liability has its own life and rationale and I regard this issue as yet another obstacle to really integrate these technologies into drug development in a rational way and something we have to deal with. But, again, I think the fear is greater than the reality but maybe I am missing something.

DR. ZACHAREWSKI: I think you have captured the fear aspect or the concern. It is a major concern and I think as the population gets balder, greater and more overweight--I am not describing myself here--you know, everybody is looking for that pill to sort of, you know, regain and capture some youth again, and you are going to find those small populations that are going to have an adverse health effect. Then they are going to go back and say, well, gene X went up and it is associated with my neurodegenerative disease and Pfizer is, you know, a deep-pocket company.

DR. WOODCOCK: Yes, from a clinical standpoint I find that somewhat implausible. I don't think from a medical-legal standpoint--I mean, we have had people who have complained that their coffee was too hot. But from a clinical standpoint we know and put on the label most of the

adverse events that are associated with a drug, the ones that are common; the ones that are even less common. It is the very rare serious ones that we may miss because they require exposure of 10,000, 20,000 people to observe one event.

Now, if you think that you are going to find that through this technique soon, I think you are wrong. But I understand that people fear that, but I think that is a very complex, probably genetic and environmental interaction usually that happens and you are not going to be able to predict that from even gene expression data.

DR. PENNIE: I think the concern that Dr. Zacharewski articulated there is more between companies having to do with plaintiffs rather than dealing with regulatory agencies, and I think it is an internal concern that organizations have to find their own path through.

DR. WOODCOCK: I agree but I think we ought to focus on what is a realistic concern. As you said earlier, some of these fears--actually, I am speaking scientifically, not as a regulator. I think you would have a robust defense usually.

DR. LEIGHTON: You briefly mentioned the problem about annotation and the difficulty this

leads to across-platform comparisons. I think this may impact on the ultimate biological interpretation of any results across platforms. Can you comment on some of the problems with annotation and a possible way forward with this problem?

DR. PENNIE: Well, one of the main problems with annotation I think, certainly for toxicology, preclinical toxicology species is, you know, incomplete genome coverage and the fact that many arrays generated in-house or even in the commercial sector, by necessity, still are not identifying a lot of the genes by name and certainly not by function. So, we have a large number of what are called expressed sequence tag identifiers on some of these microarrays which have to be continually reassessed, as more genomic information is made available in the public domain, as to whether or not those expressed sequence tags are, in fact, related to known homologs that have been encountered in other species.

So, one of the main problems, John, I think is lack of genome coverage in test species of interest. But occasionally it can also be just incorrect annotation that a particular species has

gone in 3-prime to 5-prime and so the sequence on the gene is, in fact, correct in terms of the base pairs but is completely inappropriate in terms of a hybridization experiment. So, those kind of issues we have encountered experimentally in the ILSI program where we have had a completely opposite gene expression change measured by one platform versus another and only discovered by a lot of detective work that it was an annotation error and, in fact, one of the probe sets was in the wrong orientation. So, there are many possible areas of complexity in annotation.

DR. SISTARE: Bill, I am wondering if you can give us a feel for do we need to prepare ourselves at FDA for being able to handle data on thousands of transcripts, or the concern that Tim raised earlier, is it going to drive the industry to look at known toxicants the way we are doing now to find small subsets of biomarker tandems and then just handle 10 or 20 gene transcripts at a time? If that is what we are going to see at FDA, 10 or 20 gene transcripts at a time with very focused datasets, we can do that now pretty much the way we do everything else. But if we are going to be seeing 10,000 gene transcripts submitted to us we

need to prepare ourselves for that. What is coming, from your perspective? What is going on in industry?

DR. PENNIE: Actually, that was a fairly major discussion point at the ILSI open meeting last week, and there was some discussion about the value of submitting raw data and there weren't actually very many people that were advocates of, you know, sending a 20,000 gene expression list as part of a submission in support of a mechanistic argument for risk assessment.

Again, I have to stress that as far as the ILSI committee is concerned, we are not in any way empowered nor chartered to make suggestions on regulatory policy, but it seems to me much more sensible, in a risk assessment environment, to be making a mechanistic argument to explain a preclinical tox finding and that that should stand up to a regular scientific interpretation and validation using other methodologies. In those cases you may only have to report the gene expression changes which you consider are germane to the argument you are making, but you reinforce that by using appropriate methodologies or functional work to further prove that that

mechanism is, in fact, the appropriate one.

In other words, I kind of danced around your question a little bit, Frank, but I think a combination of that kind of approach and a lot of conservatism in the industry, to me and this is my own personal opinion rather than the ILSI committee or the organization I work for, is that I suspect there is enough conservatism that you are not going to be deluged by these kind of submissions until we have a better internal comfort on the usage in a regulatory arena, and perhaps until there is a better articulation on regulatory perceptions on the state of the technology.

DR. SISTARE: All right but, given that comfort, would you foresee the future as opening of the aperture and then looking at everything in an experimental design, using a wide open array in generating that data so that you can view everything that is going on simultaneously, as opposed to looking at a light here and there?

DR. PENNIE: My personal opinion on that would be that it would be more valuable to make that information available rather than to submit it, in other words, to submit the facts which are germane, or certainly anything that is related to

the argument which you are trying to make but to maintain those records of the complete experiment locally, like we do for other methodologies; make those available for further scrutiny should the technology or the regulators desire to look at a complete dataset.

DR. SISTARE: I want to understand then what you are saying, that there would be a willingness to generate the data, to do the experiment and to measure multiple thousands of transcripts but what you are saying is the indication from industry would be to submit what they felt was germane.

That gets to the question of a lot of the same terminology that Dr. Woodcock used. Using the word "germane"--you know, these kinds of words are very difficult to define and they are moving; they are moving targets.

DR. PENNIE: Yes, yes, I agree. I agree. But that, again, was discussed at reasonable length in what I think was a very sensible and appropriate discussion that was held last week. So, I think moving forward, these issues have to be addressed really because until they are there is not going to be a significant amount of data to be quarreling

over.

DR. KAROL: Thank you very much for the presentation. Well, it is time for a break so we are going to take a 15-minute break and come back at 10:25.

[Brief recess]

DR. KAROL: I would like to start the second session with Dr. Jarnigan, who will talk to us about dealing effectively with data quality issues, platform differences and developing a database.

Topic #2 Toxicogenomic Data Quality and Database Issues Dealing Effectively with Data Quality Issues, Platform Differences and Developing a Database

DR. JARNIGAN: Well, thank you very much for the opportunity to be here today.

[Slide]

I will try to cover several of the issues that we have been discussing already this morning, particularly focusing now a little bit more specifically on what it might be that the agency might want to see as data arrives at their site. Presumably the data will arrive. I firmly believe that in time it will, maybe not today, maybe not

this year but within the next four or five years I think you will be seeing a large number of submissions with fairly large chunks of data in it.

[Slide]

Of course, the vision here, the challenge for us is that almost half of all the drugs that fail are due to efficacy and toxicology problems. Perhaps from the agency's point of view and from society's point of view and patient safety point of view, in this one-year period more than 20 million patients were exposed to drugs that were subsequently withdrawn. That is certainly a risk factor for those patients. If we could do anything to reduce those risk factors, it is a good thing.

From the industry's point of view and from the agency's point of view for better new medicines for humans one in ten INDs actually turns into and NDA. To think about that number in a different way, think about it this way, that means that all of the work that has been done, and there is a huge amount of work that is done prior to the time that a compound arrives at the agency for an IND application, you are 90 percent wrong. Nine out of ten times your predictions are incorrect. So, the vision here is to submit better compounds, safer

compounds to the agency with the belief that that will improve our odds, improve the quality of medicines that come out of the other end of the process and ultimately, because we are spending time on quality compounds, lower overall approval times.

The solution that we, at our organization, are proposing and the concepts of the agency building a database of submission data include bridging the genomic response of an organism, bridging chemistry and genomics to broadly understand a compound's effects in terms of the genomic response of the organism and, as a result of that, to have a better predictive power. That is our vision, to have a better predictive power here.

[Slide]

Before I start talking about the details of some of the features that I would think are necessary and my organization would think are necessary to make a complete submission, let me just uncover a few of the assumptions that I entered into this analysis so that the background is clear.

First off, I am assuming that the sponsor

is providing data to support an IND or and NDA application. I haven't in most of this discussion considered the fact that there may be submissions without any IND or NDA supporting feature to it but that could certainly happen. Today's discussion will focus on support of an IND or an NDA and what would be necessary.

I assume that the data is part of a larger package and is not the sole and only evidence provided to support a particular claim or a particular series of claims. That is, the data, as already alluded to, is an interlocking set of data, this data, along with other data to contribute to the claim made.

Furthermore, I assume that the sponsor has an ongoing microarray effort, and here I am limiting my discussions to gene expression microarrays, not to SNIP analysis or other kinds of genomic analysis of that kind, and if the sponsor doesn't have an ongoing effort that they will be working with a contract research organization that does have an ongoing effort. I guess what I am saying is that whatever the submitting organization, that they aren't doing a singleton experiment; that this isn't the first time they

have done the experiment; that their experimental competency in this area is large.

[Slide]

From the agency side, I also had to think about a few assumptions, and these are the assumptions that I believe the agency probably has: that the agency is willing to develop and train their staff so that the data is meaningfully interpreted and a balanced view of the interpretation is made. An over-reactive view--one oncogene is up--is not a view that would be well tolerated by the industry and not be a view that would be well tolerated by the general public because it probably would kill too many compounds moving forward.

Of course, the sponsor, and we already alluded to it in Dr. Zacharewski's comments earlier, the sponsor is concerned about about the future liability of public disclosure as well. That is certainly an issue that is in the sponsor's mind, certainly an issue that would be in the sponsor's mind going forward. I am not sure there is anything that the agency can do about this as it is more of a tort court issue but, nonetheless, it is something that has to be considered and will be

considered very carefully by the various sponsors that are submitting data.

I assume that the agency is able to accept data in a community-defined standard format and has the capability to assess its overall quality; their staff is well enough trained; their staff understands what the various features of the data are. Furthermore, it is probably the case that technologies are going to continue to develop over time and that the agency will have to continue an effort, a long-term ongoing effort to keep up with future technologies as they come forward. We are not in a static area.

The agency desires to deposit the submitted data into an internal database for use by the staff and for comparison for future evaluations, so when a new application arrives they may wish to look back at other compounds of similar type and ask have I seen this pattern before. They do this now by the use of the heads of their reviewers as integrators of this kind of data but, perhaps with electronic submission of all kinds of data becoming more and more a reality and likely to become more and more a reality, this kind of data is already set up to be electronically submitted

and probably should be so submitted.

Finally, the agency understands that the context of the data is very important, that essentially looking at a single gene or a single pair of genes perhaps isn't the best way to look at such data, and it is the pattern of the response and it is the context of that response in terms of the other data domains, the toxicological endpoints, the clinical chemistry endpoints, the histopathological endpoints that also contribute to one's understanding.

[Slide]

So, with that background, now let's talk about how array data is different and similar to traditional measurements. If we talk about a sponsor submitting a single gene or half a dozen different genes, how is that really different than the traditional endpoint?

I will just start this discussion by looking at a traditional endpoint. Let's talk about ALT elevation. It is measured. It is probably a feature of almost every IND and NDA package that is submitted to the agency. We certainly get data of that kind now. You evaluate it by looking at the mean of the groups and the

fact that no single animal within the treated group lies outside the control groups. You may conclude then that the ALT is not significantly changed by the treatment and this is consistent with good hepatotoxic toxicity. That is, it has low hepatotoxicity for the compound. So, how is that really different for gene expression data?

Now suppose that we have the case of the community, that is, the scientific community has accepted five RNAs as indicative of a certain kind of hepatotoxicity. Well, the agency and those companies may well get data of the following kind wherein they have the five genes measured as the ratio to control, for example. They have the means and the standard errors. They know that no single individual treatment was outside the range of the control. Would it be reasonable then to assume that these RNAs are not changed? The answer is probably yes. So, again, the sponsor might conclude that there is no significant change and it is consistent with good liver toxicity, that is, low liver toxicity.

[Slide]

But microarray is different from conventional measurements in some ways, the first

of which is that both the agency and the community have a lower familiarity with the technology. It is new technology. There are features that are different from traditional measurements. Of course, this will improve over time. Five years from now this discussion probably will be much, much less significant.

There is concern that the survey nature of the data might uncover confounding factors, factors that the sponsor would rather not know about or that perhaps could be confounding to an interpretation. The sponsor, of course, is concerned by an overly reactive view. A certain gene has changed, therefore, we can't go forward. That may be overly reactive.

Of course, the agency perhaps has a concern that the sponsor is missing important findings, remembering that the agency may well get data arriving at their site from a new therapeutic class never before exposed to patients but this is the fourth application in the last two years they have seen. They may understand things that the sponsor even doesn't understand. I already know that the agency gives Greenspandian kinds of comments where they say, "we think that you ought

to look at the kidney" as a statement. Of course, you have to react to that even though you don't understand why it is important that that be done now.

Finally, I think it is very important to note that there is less scientific agreement about how to interpret these findings. This is an area, as Bill Pennie mentioned, of logarithmic growth. The methods for interpretation, the way you go about these kinds of interpretations are improving logarithmically right now. Pattern matching is a key component of this, and this is less familiar to the biological community. We are used to looking at a single group of genes, a single endpoint. So, it is an unusual treatment of the data for most of us. Furthermore, it is different than most of our training as we came along through our various educational paths. It is going to take some time for the community to be educated about this kind of an approach, but it will happen. It will happen faster than we think. I think it is penetrating already and will happen even more quickly than we think.

Finally, I would like to point out that there is a perception that microarray data is lower

quality and noisier than our traditional measurements. Certainly, five years ago or four years ago that was a very true statement. Today the technology has improved dramatically. The quality of this data is getting to be very high and, when competently executed, I believe it is approaching the quality now of almost any other traditional endpoint and in another five years I think it will be there. So, carefully conducted experiments are accurate and predictive, and they will get even more so over the next several years so this issue should slowly diminish.

[Slide]

Now let me just summarize what I think a sponsor might want to provide to the FDA in terms of a package of information for microarray data, then we will go through each of the points more or less one at a time. I definitely would urge that the sponsor provide MIAME or MAGE-ML compliant descriptions of experiments and electronic submission of all data. It is not useful in this context to submit data on paper--10,000 measurements at a time, 50 microarrays in a typical submission perhaps. It is just not useful.

Minimum experimental design metrics

similar to that required for any other biological experiments are a definite must. Four or five years ago you could definitely find papers in the literature where a single microarray comprised the whole publication. It was the case where scientists said, well, I am measuring 10,000 endpoints so I don't need to do triplicates; I don't need to do multiple biological controls. That is just not acceptable and shouldn't be acceptable here. I don't need to tell the agency how to evaluate biological data, they do it every day, but we need to remind ourselves that that is important.

The novelty of this technology requires that additional quality data be submitted to demonstrate the competency of the experimenter. That is true for today and for the next several years. Perhaps in time we won't be questioning the competency of our experimenters but for the next few years I certainly think that that is a probable, definite thing that will have to be done.

I would definitely urge the sponsor to provide and interpret the data in a scientific style format. That way the reviewers, particularly in the IND setting where they have only 30 days,

don't spend tons and tons of time digging through mountains of data. They can go to the paper, read it and then, if they have further questions, they can dig again to a specific point.

Finally, it is very important, we found at our organization, to compare to community accepted RNA biomarkers and comparing to bench mark drugs and toxicants is extremely valuable. It provides the kind of context that you can't get through other approaches. So, the interpretation needs to be in the context of current drugs, failed drugs and toxicants. I think that is a very important feature.

[Slide]

In the next minute or two I will talk about these minimal standards, a little bit about the quality control data and something about this scientific interpretation. So, in the next few minutes the themes that I am going to delve into with the quality control are constant. There will be three of four different kinds of endpoints that I suggest but their themes are fairly constant.

First, measurements versus the lab historical values. Again, my assumption is that a lab is running these experiments all the time and

could easily generate the historical data that is necessary by which to compare the quality.

The measurements versus an external standard--the agency and NIST are combining to try to define a standard. Definitely, we ought to be carrying these standards through with any experiment that is to be submitted. To provide that data and measurements versus the external standard will be very important.

Measurements versus an internal standard. All manufacturers that I am aware of provide a certain number of spike-in standards to include. You ought to use a few of those and include that information as part of your quality control measurements.

This is a little bit different than a traditional submission to the FDA and that is, of course, because of the youth or novelty of this technology. You have to prove your competence at doing the experiment and you need to assure the competency of the experiment or you need to assure that it is consistent with internal and external standards and need to assure that it is consistent with historical values. All of those things should be possible in almost any laboratory that is doing

these studies routinely.

[Slide]

Now, the experiment to create a microarray finding from a drug-treated animal is actually a fairly complex experiment. By our count there are 286 steps going from a drug in a bottle to a finished microarray experiment at the other end of the process.

This pattern is similar for all the different platforms. You do an in vivo experiment. You isolate the RNA and you prepare a target of some sort. You hybridize that. You check the quality of your final product and you load it into an array. Most labs will have some sort of a minimal laboratory information management system underlying this data generation process. So, generating this historical data comparison to controls, and what-not, shouldn't be a big problem.

But there are three or four points during this process where I feel it would be very important that minimal information be collected to, one, prove the competency of the lab doing the experiment and, two, to assure anybody else looking at the data now or five years from now or ten years from now that the experiment was done well. Those

are shown at the end of the in vivo experiment, the end of the RNA preparation and then at two or three different kinds of checks relating to the quality of the hybridization. These points I believe are independent of platform, and very similar numbers could be found for all different platforms.

[Slide]

First off, just let me mention a few words about the minimum experimental design just to remind everybody that the minimal experimental design, at least in my mind, is that you have at least three treated samples; you have at least three control samples; and that you carry through with your process contemporaneously three of these RNA standards, external RNA standards, as well as carrying through all samples three spike-in RNAs as a minimum. This would then impute that the minimum experimental size to be submitted is nine microarrays with three RNA standards in every sample. So, minimum biological triplicate; minimum of three untreated or mock treated vehicle controls, processed contemporaneously with the samples to be run; a minimum of three external standard RNAs, also processed contemporaneously with the samples under consideration; and a minimum

of three spike-in RNAs.

[Slide]

Now moving on to the RNA that is used in the experiment, there are a number of different procedures for preparing RNA but they all end up with a product that contains 28S and 18S RNA. They are present in all samples. I propose that the community settle that at the very minimum the mean and the standard deviation and the range for the 28S and 18S RNA, the amount of that and the ratio, be reported and probably the traces for those various RNAs that support the package of data be provided. That way, ten years from now if some retrospective analysis is going on and you wish to understand this material the data is available. It is not too much to ask most of the labs. They all have this information in electronic format today so adding it to the data package is not that difficult.

I propose that this data be provided for the samples in the dataset for historically similar tissues or cells prepared in that lab, again testifying to the lab's consistency and quality over time, and that the data be provided for this external RNA sample that is executed or processed

contemporaneously with the data.

[Slide]

Now moving on to the hybridization, quality control for the hybridization, there will be two different kinds. First, I propose that for every microarray that is run that the array average signal to background ratio be computed; the array average background; the average raw signal; the log dynamic range for the signal; and the average signal intensity for the three spike-in RNAs, minimum of three spike-in RNAs be reported, and it be reported in some sort of a data table that compares it to historically similar samples for matched tissue type or cell type being run in the lab; the historical samples averaged for the RNA standard that is being run; the historical average for the spike-in RNAs; for the contemporaneous RNAs; and for the contemporaneously run standard.

With that, one can easily look at the data and say it is very consistent and this lab can execute a consistent experiment over a long period of time. Again, I am assuming that the lab is processing samples on a fairly routine basis and has this information available to them.

[Slide]

The last point I would like to make about the quality of the experiment has to do with the internal and external consistency of the samples. One of the easiest ways to measure this is to measure the correlation coefficient for any pair of samples in your dataset. Just assuming three, then you have two pairs in your dataset and you can measure the correlation coefficient versus each other; versus the contemporaneous control; versus the contemporaneous external RNA standard; perhaps versus a historical RNA standard, again getting back to the fact that the lab can do the experiment consistently; and to historically similar tissues or cell types. The report then for the dataset provides the mean and the standard deviation, and perhaps the range of the correlation coefficients for those various datasets.

[Slide]

That then concludes the main quality control points that I would suggest be included in a submission. Now turning my attention for just a minute to what might be submitted as an interpretation of the findings by the sponsor, I think that should be somewhat in scientific literature style format. That means it starts with

an abstract, remembering that, particularly at the IND stage, the reviewer has 30 days so they don't have an infinite amount of time to review this information. They need an abstract; something about the significance of the experiment relative to the specific application under consideration; a brief methods because somewhere in that MIAME submission there is a very long and detailed methods and it is not necessary to make the reviewer wade through that to understand what was done but a brief methods should be provided here; a summary of the quality evidence described earlier; something about the results and a discussion of the results; then conclusions relative to the specific application under consideration and conclusions in the context of a wide variety of other drugs, standard toxicants and failed drugs that are available on the market, that is, some sort of comparison to an external database of some sort. Of course, by providing this summary of the results you are helping the agency help you. You are helping them direct their attention to important points in your data and providing them with some understanding as you see the data.

[Slide]

So, in summary, I propose that MIAME or MAGE-ML compliant descriptions be provided; a minimum experimental design metrics similar to that you would do for any other kind of a biological experiment. Let's not treat this any differently than other biological experiments. For the next few years at least we need to provide additional evidence that the lab is competent to perform the experiment. Perhaps in time that will go away but today we need that. Your interpretation of the findings, and then a comparison to community accepted RNA biomarkers, so appealing to whatever is in the literature, and comparison to bench mark drugs and toxicants. Your interpretation should look outside the dataset provided.

[Slide]

Now let me talk a little bit about this external dataset and how one might go about the comparison, and also talk about how the agency might want to build the database comprised of the submissions as they come along, with the goal that in time they will have a contextual view of new submissions as well as a contextual view to look at for things that are approved, close-failed relatives in certain standards and toxicants.

It is my belief that the agency might want to build a contextual database. Microarray technology will require that we step into the coming age of electronic submissions. We are still getting a lot of submissions, I understand, at the agency that are largely paper in nature but we will be going into electronic submission and microarray data is already electronic in format so it can probably lead the charge here. Paper submission of microarray data is not very useful. If you think of a million data points on paper, it just doesn't provide any interpretive context for anybody. The agency is probably not going to retype that data into a computer to analyze it so it has to be done.

I believe that this contextual database will be used by the agency to better understand the technology. It will be used by the agency to look at the data in the context of other submissions, remembering that the agency may well get data and have a view on data that is not available to the sponsor because new therapeutic modalities are being presented to the agency that have never before come along. So, they may have a view on data from two or three of these that the rest of the industry doesn't have. The contextual

database, in our experience, is highly useful to provide meaning and a balance to the interpretation, and I would like to illustrate the point about the balance in a slide or two.

[Slide]

Before I do that though, I would like to turn my attention to what will the agency do with this data. Again, promoting a balanced view has got to be one of the central objectives. It is very easy to overreact to some single data point or two or three in the data. You need to be aware of what truly significant events are. The way you get that awareness is by developing a community consensus around what are useful RNA biomarkers, and the way we get that community consensus is by doing a lot of experiments. So, you need to ground the analysis in the context of real-world effects of drugs, failed drugs, withdrawn drugs, standards and toxicants. So, a reference database is needed.

[Slide]

Such reference databases are being produced and prepared now and are available. What should be in one of these reference databases? Well, it should contain a wide diversity of successful drugs, failed drugs, toxicants and

standards. That is, you need to understand both the pharmacology of compounds as well as their toxicology. In our experience one cannot truly divorce those two fields, one from another. You must understand what the drug does pharmacologically as well as toxicologically.

The database probably should include multiple tissues, doses and times, and probably cells in culture as well. The linkage of the expression data to orthogonal data domains is very important. You find a lot of good, useful new insights by understanding what goes on pharmacologically, including site interactions with on and off target events. What happens with the histopathology in animals dosed with these compounds, clinical chemistry, hematology and chemical structure are all useful orthogonal data domains and should be present in a contextual database, and in vivo and in vitro experiments so that you may bridge between your in vitro findings to your in vivo findings.

[Slide]

Let's just look at what the benefits of using a reference database are. We have heard allusion to this kind of result both in Janet's

talk and in Bill's talk earlier. This is data taken directly from such a database looking at three oncogenes. I just picked out three to look at them, just for illustration, EGF-receptor, cKit-oncogene and BCL2. All of these drugs cause statistically significant elevations of these oncogenes.

One single oncogene change is certainly not significant. It is certainly the case that these oncogenes, as Janet says, weren't put into the genome to cause cancer; they are there for the cell and the organ to respond to specific environmental stimuli. Drugs are environmental stimuli and they, therefore, cause changes in these oncogenes. Elevation of one is not in itself evidence of cancer. These drugs are not oncogenic in general.

So, the context provided by such a database provides a balanced view and will accelerate the adoption of this technology because we won't have to wait for these experiments to be done as singletons in individual academic labs over the next several years.

[Slide]

So, to summarize and then move on to

looking forward, electronic submission of the data--a definite yes. Standard format--a definite yes. Perhaps the agency should help the process by helping devise some sort of input tool for the standard data format, a better input tool than is currently available. I am reminded very much of what it was like to submit data to GEN Bank before SCAN was available. It took hours and hours just to get it into the form to be put into GEN Bank. Once the SCAN tool was provided to the community it went much faster. An analogous situation happened with PDB a few years before that where data was submitted in all sorts of formats. It was impossible to database. Once an input tool was developed and Brookhaven took over the job of putting together a simple database it became a useful tool.

Minimum experimental design--we can't forget what we learned on how to design biological experiments years ago. It is still valid in this technology. New technology does not obviate those needs.

For the next few years, perhaps diminishing with time but for the next few years the experimenter needs to prove their competency at

doing the experiment by providing additional data beyond what would normally be provided with any other kid of biological endpoint.

Sponsor's interpretation of the data I think is extremely important. It should not be ignored. A pile of data should not be submitted without much support as a written document of some sort.

Finally, comparison to community accepted RNA biomarkers, there are some in the literature already and we should definitely look at those, and also comparison to bench mark drugs and toxicants, withdrawn drugs and so forth.

[Slide]

So, conclusions and looking forward. Microarray technology is ready to contribute to the drug discovery process and to the approval process today and I believe that as we start to do this we will start to see improvements in our overall efficacy of this process, improvements in the safety of compounds that are submitted, improvements, therefore, in the overall quality of medicines that are being used to treat patients.

Simple assurances of quality are definitely needed for the time being. Contextual

databases to allow meaningful interpretation are needed and some are available. We need to develop as a community a consensus around what are meaningful RNA markers. This is starting to happen. I think it will accelerate over the next several years.

Again, requirements beyond normal verification of data quality will diminish as community sophistication improves. I will say we have done a number of experiments analyzing data collected over different platforms that can make accurate predictions on data prepared in several different platforms. The same biology is found regardless. These technologies all do measure the same biology and that is the critical event. That is what we are after, to measure the biology and understand that that biology is significant for safety or for efficacy.

Finally, I believe and definitely know that clinical applications in accessible human tissues for this kind of RNA transcription measurements will come and will be parts of submissions very shortly to the agency.

[Slide]

So, the result of this activity--building

a database, providing the data in an electronic format carefully controlled--will be to improve the predictive power of the animal studies that are undertaken and of looking at clinical samples in accessible tissues. This will help realize this vision to get better compounds submitted; safer compounds submitted and approved; and lower the overall approval time because we spend our time on the best compounds. Therefore, we are addressing the problems of patient exposure to drugs which are subsequently withdrawn because there are fewer subsequent withdrawals perhaps. It addresses the problem that only one compound in ten enters and IND passes an NDA test. Thank you and I will be happy to take questions.

DR. KAROL: Thank you very much. We have time for perhaps one or two questions.

DR. GOODMAN: I like the portion of your presentation dealing with providing the information in the format of a scientific interpretation. But just to be a little argumentative, why do we need the rest? That is, it seems to me that one way that would stifle what I think is a very promising technology is to, at the outset, be too prescriptive as to these are the way the data will

be submitted; these are the types of information that one wants; and maybe also to be too prescriptive in terms of talking about setting up a database if it will result then in driving, if you will, the experiments. That is, now the data must be submitted to fit the database as opposed to what scientifically might be best.

DR. JARNIGAN: First off, I would point out that if you read the MIAME and MAGE-ML standards, they actually have a tremendous amount of latitude built into them. They aren't overly prescriptive. Perhaps I am wrong but certainly I don't read them as being overly prescriptive. Provision of the data as a whole, meaning all 10,000 genes or 20,000 genes at a time, that is an issue that, as we discussed, will be difficult for the community to address and I think the difficulty isn't with the agency; the agency can handle this problem well. The problem is the tort issue. The tort issue probably has the pharmaceutical companies more concerned. So, they are worried about the future liability--the issue that was brought up over here earlier today--the future liability for something being discovered five years from now or ten years from now that says you should

have found this ten years ago. We don't proscribe it on ourselves now. I certainly know that submissions arrive that have issues that ten years from now are bound to be a problem but, still, it is going to be something that they consider very heavily.

To your question, I think that your question is are we proscribing it too much? Will this make the experiments fit into a nice, neat box? I don't think the electronic submission standards do demand a nice, neat box. They just demand certain basic things, many of them you already require of yourself for all other kinds of data that you submit to the agency.

DR. KAROL: Thank you. I am afraid we will have to move on. Thanks very much. The next presentation is by Dr. Quackenbush on data processing, statistics and data presentation.

Data Processing, Statistics and Data Presentation

DR. QUACKENBUSH: Thank you very much for the invitation to come here.

[Slide]

My background isn't in toxicology; my background really is in other areas of applications for microarrays so I may not be able to address all

the questions specifically associated with toxicology. What I am going to try to do is address questions associated with data handling and management and, as Frank asked me to do, try to point out what some of the issues and challenges are and take you, if I have time at the end, through one or two examples where we have tried to apply some of the lessons we have learned for understanding array data.

I have prepared a handout for you and I have already deleted a large number of those slides. I tend to have too many slides always and am then deleting them in the last few minutes, but I haven't rearranged the order so you won't have to skip through too much.

[Slide]

What I really wanted to start with in looking at this problem is actually just looking at the problem from the start, which is selecting the appropriate platform.

[Slide]

This, in fact, can be a bit of a challenge. As you know, there are two array platforms. One is a resequencing-based platform that developed out of the Affymetrix resequencing

chip in which oligos are synthesized de novo on a glass substrate.

[Slide]

Then two biological samples are labeled, hybridized independent arrays, scanned, relative expression levels are measured, and from that relative expression level measurement on two independent arrays one can derive changes between a query and control sample or between any two samples in the experiment.

[Slide]

The alternative approach is to take DNA fragments, whether PCR products or long oligonucleotides, and array those on a glass microscope slide using a robotic spotting device, and then RNA is extracted from two different samples. In this case, the RNA is labeled with distinguishable fluorescent dyes, although that is not always the case. Some people treat these arrays also as single color assays and perform independent hybridizations, but the most common implementation, in fact, is to use these paired samples, hybridize them to a single array; measure fluorescence intensities and analyze them to identify patterns of expression. The real

challenge, of course, is to take those patterns of expression and interpret them in some kind of meaningful biological context.

[Slide]

This was supposed to unfold and it really didn't unfold very well at all. Somehow it got rearranged in transfer. But, fundamentally, the array assays start with looking at genes because that is the object we want to understand. Those are represented by one or more elements on the array. We measure fluorescence intensity for each one of these elements and from that an inferred expression. We like to link that back to the gene.

In fact, every part in this process has potential pitfalls and is problematic. One of the most important is moving from spots on the array to relative expression measurements. This is something which I know was discussed to a certain extent this morning but it is absolutely important. All of the laboratory handling of the samples--how you choose the samples; how you deal with them--has a big effect on what you ultimately measure. In fact, we are not measuring expression, we are inferring expression based on fluorescence intensity, which is based on hybridization, which

is based on relative RNA levels. So, if the samples are allowed to degrade at room temperature for a long time before the RNA is extracted, if the RNA is degraded before it is labeled, then what you see on the array expression may or may not, in fact, really be the relative expression for those genes.

The other important aspect is that what we call the genes on the arrays really have to be carefully defined because those genes, in fact, may not be what we think they are when we look at the annotated elements on the array. I will come back to one or two sources of that in a minute.

[Slide]

So, there are some platform related issues. One is the lack of standardization which makes direct comparisons of results between laboratories a challenge, not an insurmountable challenge but definitely a challenge.

This says "lot-to-log," in fact, it should say lot-to-lot variation in arrays. Lot-to-lot variation in arrays can introduce artifacts and the results can be dependent on either the biology or on artifacts on the arrays, and that can include the log-to-log variation as well as which

technician performed the assay, which day of the week they did it, the reagent lot. So, all of those have to be very carefully managed and controlled to make sure that when you are actually looking at an experiment what you are seeing is the real variation that comes from the biology, not from the fact that the arrays were done on Wednesday rather than Friday when everybody was ready to go home.

Commercial arrays provide a standard and remove some of the design considerations, in particular the idea of using one sample per array which makes all of the experimental design much easier. It presents different challenges for doing analysis, but the cost is significantly greater for doing these commercial arrays or using these commercial platforms which drives a lot of array users, particularly academic users, to use in-house arrays.

But no matter what, one of the most important things, which I tried to emphasize earlier, is really the demand for a good LIMS system to track every single aspect of the experiment. Those have to be tracked not only to report them but, in fact, to really interpret and

understand what you are seeing and to identify potential sources of artifacts.

[Slide]

Once an array platform is selected we want to move on and actually start doing array analysis.

[Slide]

There is a general strategy for doing the microarray analysis. The first is to choose an experimentally interesting and tractable model system. To design an experiment with comparisons between the appropriate variants and to include the appropriate controls you have to include sufficient biological replication to make good estimates, which is a point that has been emphasized here before. Once you have designed the experiment and start doing hybridizations and collect data, that data has to be effectively managed. The data then has to be normalized and filtered so you can make appropriate comparisons between different hybridizations, different individuals, different labs, different experimental protocols.

Then, and only then can you begin to mine data to look for biologically interesting patterns of expression. Then, in order to interpret those patterns of expression, you would like to integrate

the expression data with other ancillary data, including information like the genotype, the phenotype, the genome, the annotation of the genome, the treatments you are using, the dose, the dose response, other physiological measures. In fact, probably the biggest challenge is moving from looking for these patterns of expression to really trying to interpret what they mean based on the underlying biology.

[Slide]

The first step in doing all of the data analysis is actually having useful annotation on the array.

[Slide]

While this may not sound like a significant challenge, in fact it is. You may have read that the genome has been finished yet again, the human genome. That was published in April of this year. Based on my definition of "finished"--that we have a complete genome sequence; that we understand where all the genes are; we have functional assignments for those--the genome is far from complete. That doesn't mean that the draft human, mouse and rat genomes are not useful. In fact, they are tremendously useful for

analyzing the data. But one thing I want to emphasize is that they have to be taken with a grain of salt.

So, we do annotation on the arrays that we build in-house and for the array assays we perform in-house. These are built around a series of databases we call the TIGR gene index databases. I am going to talk about these databases only because for us the annotation process is important in understanding potential pathologies that arise in that annotation, important for interpreting the results.

[Slide]

So, we have built these now for nearly 60 species. This is an example of what one of those records look like. It comes from taking gene and EST sequences. ESTs are still important even in the realm of the complete genome because many arrays have ESTs representing, including a lot of the commercial arrays. So, we take the ESTs and gene sequences. We assemble them. We provide information about those assemblies, links to public databases and information such as annotation based on sequence similarity search and gene content, links to other databases, in this case to the mouse

genome Informatics database at Jackson Labs, and increasingly maps of things like the completed genomes.

[Slide]

Another important element of the annotation though is to try to understand the functional roles that these genes play and, in particular, for interpreting the results in the context of the biology you are examining, being able to project additional annotation and classification ontologies onto the genes is incredibly important.

So, one of the things we use are the gene ontology terms or GO terms. Gene ontology is an attempt to define in a rigorous fashion classes for genes in three broad categories. The first is molecular function; the second is biological process; and the third is cellular component. So, what we try to do is take each one of our array elements and attach this kind of annotation which allows us to place genes in broad biological classes.

An additional attempt that we make in annotating our array elements is to provide EC numbers. The enzyme commission numbers allow the

array information to be projected back onto things like metabolic pathways.

[Slide]

We are also very interested in building cross-species comparison. We built a database which is known as EGO, the eukaryotic gene orthologues.

[Slide]

What this database attempts to do is to use pair-wise comparisons between sequences to identify possible orthologues requiring transitive reciprocal best matches between multiple species in order to define an orthologue set.

[Slide]

This has actually been very useful for identifying orthologues in mammals as well as across kingdoms. So, in this case what we have are sort of orthologues from human, mouse, rat, zebra, fish, potato, tomato, barley, beet, rice, maize. In fact, even using DNA sequencing you can identify these.

In the context of toxicology, while looking at human or arabidopsis orthologues might not be that interesting, really identifying the human, rat or mouse orthologues is going to be

fundamental for interpreting a lot of the data.

[Slide]

One of the other important lessons I think we have learned in looking at this data is just the value of seriously questioning the annotation that is provided for the genome sequence, and these are just some examples I would like to show. These are the official ensemble gene predictions, as well as alignments to EST data from human, mouse, rat, cattle and pig, the most highly sampled mammals.

In many instances the ensemble annotation is quite good and recapitulates the gene structures that you see in these other species. In other cases there are ensemble annotations which have no EST support despite having nearly 15 million mammalian ESTs available. There are other very clear examples where there is beautiful EST support among multiple species or a single species but no annotation.

So, one important lesson to learn is that the genome and its annotation is only a hypothesis. That hypothesis still remains to be tested. In fact, one of the things I didn't emphasize at all is that the assignment of gene function to many of these genes is based only on sequence similarity,

and sequence similarity search is not an actual experimental evidence.

We have many good examples, in particular for arabidopsis where there has been a complete genome duplication, where genes that have been assigned exactly the same function in fact respond very differently and have clearly different functions. The annotation is an ongoing process in biological interpretation of response to any kind of challenge using array data and it is really going to require careful follow-up of what that annotation is.

[Slide]

Another important aspect of this entire problem is to try to address this cross-species comparison and the cross-platform comparison problem.

[Slide]

In order to do this my group built another tool, that we call Resourcerer, that allows you to take microarray resources and provide annotation for them, including things like links to locus link, links to the physical map and orthologue identifications and gene ontology assignments.

[Slide]

This tool, based on having an orthologue database, allows us to compute cross-species and cross-platform comparisons so in this case it is a cDNA clone set linked to the Affymetrix human U95A array. Another important element is having access to the genome sequence, in which case we can take things like genetic markers and simply ask questions, if we have an area of the genome that has been linked to a particular response through genetic mapping, can we find elements on the array that will allow us to provide an intersection between genetic data and expression data.

In the context of testing compounds this may not be important; in the context of understanding response it may be very important as different mouse and rat strains, in fact, are known to respond differently to different challenges.

[Slide]

So there are real annotation issues. The first is the complete genome is incomplete. The gene names are not well defined so one gene may have many names. One gene may have many sequences representing that gene and they may not be the same sequences, and one sequence, in fact, may have many names. So, looking across the aliases for each

gene can really be an important problem and this is one place where standardization can be absolutely essential and helpful in interpreting results.

Analysis interpretation depends on having well annotated array elements and gene sets, including gene names, gene ontology assignments and information about pathways. Cross-species comparisons also require a very careful analysis and knowledge of orthologues and paralogues in order to draw the correct inferences.

[Slide]

Another important area in terms of applications and annotation and analysis is developing appropriate tools and techniques for analysis.

[Slide]

I am actually going to skip a number of the slides I put in here, which is sort of elementary introduction to some of the challenges, but there are important steps in the entire analysis process.

[Slide]

The first is choosing an appropriate experimental design. In fact, in the statistics community, as you probably know, there has been a

great deal of discussion and debate about what the appropriate experimental design is and I can tell you that there are important differences between statistically sound designs and experimentally tractable designs that aren't always addressed in these debates in the literature. So, those have to be addressed appropriately and carefully.

You perform the hybridization and generate images. You analyze these images to identify genes that are differentially expressed and their expression levels, usually measured as hybridization intensities. The data is typically normalized in a variety of different ways to facilitate comparisons between elements on a single array and between multiple hybridizations, and then we want to analyze the data to find the biologically relevant patterns of expression.

[Slide]

Again, I will just mention that my group builds a lot of software for addressing these issues and if you would like to talk about particular algorithms we can discuss them.

[Slide]

The first piece of software I showed you is actually our data management software that

allows us to track information through the lab. All this software we provide to the community with source code.

[Slide]

One step in the process though which is absolutely fundamental is normalizing expression data. Normalization is actually important for facilitating comparisons across arrays. One of the simplest things you can do is to simply look self versus self hybridization, compare a hybridization assay to itself using either a two-color assay or using multiple hybridizations across multiple chips with the same sample.

What you would expect in an assay like that is that every gene, in fact, should give you a ratio of one or a log ratio of zero. In fact, you know that is not true. There may be unequal labeling efficiencies or hybridization or detection efficiencies for the different dyes. There is, in fact, inherent noise in any measurement you make and there is noise in the systems that are used. In fact, even when we are looking at self versus self hybridizations comparing the same sample to itself, we may, in fact, be seeing biologically relevant differential expression if we are taking

two RNA extractions from the cell line drawn in two different flasks in the same incubator. Not all RNA is equal and handling those samples can affect them.

So, very often when people look at this kind of self versus self hybridization they are not seeing what they expect because they are not looking at what they expect. Normalization is a process designed to bring appropriate ratios back to one.

[Slide]

The technique that we use for looking at two-color microarray assays is locally weighted linear regression in which we try to subtract out this sort of systematic curvature you see. What we are looking at is the logarithm of the ratio. It is really a measure of the log of the intensity on the array, and we try to center that data and also smooth it out. Whether doing that centering is appropriate or not is, in fact, open to interpretation and really depends on what the biological experiment is that is under way. Probably the nicest discussion of this is a recent paper that appeared from Frank Holstege and his group in which they looked at a situation in which

transcription is shut down and normalization of the data, as it is typically performed, is not appropriate.

One of the other things that is important to realize is that when people talk about differential expression, how they actually measure that differential expression is fundamental to interpreting the result and often ignores the real structure in the data. So, if we look at the log to the ratio and, in fact, pick a two-fold up or down regulation, two-fold here is represented as a log ratio of plus one or minus one. In fact, at low intensities, as we approach the detection threshold on the array, two-fold may be completely meaningless, while at higher intensity something like 1.2- or 1.3-fold may, in fact, be a significant change. So, we have to be very careful and very intelligent about the way in which we even identify what we mean by differential expression, and we have to use the appropriate tools for identifying genes, including the appropriate statistical tools.

[Slide]

Again, my group builds software for doing some of this normalization, as well as doing data

analysis and we can talk about the various algorithms.

[Slide]

There are some issues though. The first is that there is no standard method for data analysis. In part, that is tied to the fact that there is no standard method for experimental design. The same algorithm with a small change in parameter, such as a different distance method, can produce very different results when we are analyzing expression data. Data normalization plays a big role in identifying the differential expressed genes and how you scale within and between arrays can affect the results. Much of the apparent disparity though that is observed in microarray datasets, in fact, can be attributed to differences in data analysis methods. When people pick out a group of genes from one set of experiments and do experiments on a different platform and pick out a different set of genes and they say, oh my God, they are discordant. In fact, that may not be the appropriate test because how you pick out that class of genes depends on the assumptions, depends on the software, depends on the parameters. In fact, my analysis and the

analysis my group has done seems to suggest that a lot of that comes from the different analysis methods, starting with things like image processing and moving on to normalization and data mining.

[Slide]

Another important element which has been discussed here at length is data reporting standards so I am not going to discuss this in very much detail, other than to say that I have been involved in this MIAME consortium to try to define standards. Really, the emerging standards are that we have to report everything that is relevant to the measurements that are made on the arrays.

[Slide]

The good thing I think which is motivating the community to adopt these standards is that the journals themselves have been asking for the standards to be advanced and now most of the large, high profile journals require that data be submitted in a MIAME compliant fashion.

[Slide]

One of the important things I think that is emerging from all of this is the development of an extension of MIAME called MIAME-TOX. If you want to take a look at this standard, it is going

to be discussed in greater detail at the upcoming MGED meeting in September, in France. But, clearly, implementation of all these standards is going to require development of ontologies to describe the experiments in more detail, the analysis tools in more detail and, in fact, the experimental challenges, particularly the toxicological challenges in very clear, well-defined detail.

[Slide]

Our software also has to be developed to read and write MAGE-ML. There was a question about the flexibility of sort of the openness of MIAME and MAGE-ML. MIAME in fact was initially proposed as a very flexible standard, in large part because I think we realized within the community that the standard is still being developed. In a similar fashion to the MAGE-ML, the XML-based reporting standard is very open to development of new applications and new techniques in particular extensions which will be appropriate to toxicology.

[Slide]

The public databases clearly need to be extended to meet the toxicological needs or new databases have to be created to include that

information.

[Slide]

I wanted to talk a little bit about some of the science. In fact, what I am going to do is I am going to skip a lot of this talking about the biology, but I am going to bring up one important issue.

[Slide]

The two examples I was going to show you are an example of how we use genetic maps to try to refine expression data; another one in which we use GO terms to try to refine expression data.

[Slide]

One of the things I am going to talk about very quickly is the problem of trying to predict outcome since that seems to be a lot of the challenge in toxicology. The problem for us is that we are looking at patient samples in a cancer study funded by the NCI in which we want to try to use expression fingerprints as a phenotypic measure for predicting things like survival, response to chemotherapy and outcome.

[Slide]

The first problem we wanted to attempt to address is a problem which is very simple, the

problem of classifying tumors. So, what we did is we took a number of adenocarcinomas. We profiled them on 32,000 element human arrays.

[Slide]

And, we used a variety of techniques for predicting which genes would, in fact, be the most appropriate for classification. The approach we finally chose was one in which we used the neural network and in terms of toxicology, neural networks may in fact be problematic because they are black boxes. In terms of doing classification though they are actually quite effective because what we can do is use input data, and here the input data are statistically significant genes which are good for separating out different tumor types and now can be trained to predict the class of tumor.

[Slide]

We built a classifier that was 94 percent accurate using data on cDNA arrays. Part of the reason I wanted to talk about this experiment at least a little bit is because what we realized we needed to be able to do is to extend this classifier. So, we surveyed the literature and found available data that we felt we could use. For a variety of reasons, the only available data

that was published that we felt we could use was data that was collected on Affymetrix chips.

[Slide]

So, we scoured web sites. We downloaded the data. We ended up with 540 tumor samples representing about 95 percent of all human cancers, representing 21 different tumor types.

[Slide]

The real challenge, of course, was to be able to do a cross-platform comparison in which we were really looking at three platforms because even the two Affymetrix platforms don't have the same probe sets for all of the genes on the array. If you have the same gene you may, in fact, have two different probe sets.

So, we had to do some kind of cross-platform normalization. The approach we used for this was actually fairly simple. On our spotted arrays we compare everything to the universal reference. What we did was we took these Affymetrix arrays and we hybridized our universal reference to those arrays and used the data on a gene by gene basis to scale each one of the expression levels. Having done that, we got a dataset that was comparable that we could then use

to train this classifier and actually make tumor predictions.

[Slide]

The short version of this is that at the end of the day, even looking across multiple platforms, we were able to build a classifier that was nearly 90 percent accurate, approaching the level at which a pathologist, over the course of a number of tests, can actually classify these same tumors. We have extended this now to look at survival and to predicting outcome, and I can tell you that it has been equally successful in these other applications.

[Slide]

So, what are the real challenges in analyzing microarray data? One is that statistical significance is not necessarily the same as biological significance. Having enough replicates to define statistically significant results is important but it is not the only thing, and one of the things we have to remember when we analyze this data is to look at the biology.

Another real challenge which I think people are realizing is that if you take this system and perturb it many genes change their

expression levels, not just one. So, in fact, a very simple challenge in which you try to just perturb one single pathway can produce a lot of unexpected changes, and those changes may be difficult to understand. One of the first observations we made in tumors is that genes like osteoparten change. We reported this in a paper and one of the referees wrote back and said obviously this data is nonsense because osteoparten is a bone protein. So, really you have to be very careful at how you look at these and how you interpret the data in light of the annotation.

Multiple pathways and features in the data can be revealed through different analysis methods so the same dataset can show you four or five different patterns, depending on how you look at it and how you interpret it has to depend on biology.

Genes which are good for classification or prognostics may, in fact, not be biologically relevant in the sense that there may be some of these ancillary changes that occur as you perturb the system, and they may be very important for making the predictions but they may not tell us about the biology.

Finally, extracting meaning from

microarrays will require now software and new tools, but the most important thing we need is more data collected and stored in a standardized fashion.

[Slide]

I am seeing that I am running over time. The most important thing I think really to take out of all of this is that there is still a lot of need for standardization but one of the most important needs we have in terms of developing statistical tools and analysis tools and techniques is just good data which is collected and stored in a standard way.

So, thank you for the invitation and thank you very much for the opportunity to talk here today.

DR. KAROL: I would like to take just one short question.

DR. WATERS: I think you accurately captured the complexity of this field that we are evaluating today. The question that I have, and really in a way it is a comment, has to do with the capture of the toxicology side of the dataset. You mentioned that briefly as you went through the evaluation of the various types of measurements

that should be made. Could you comment a bit more about what you really think the importance is in capturing that data. We heard in the previous presentation that context was all important but we didn't hear anything about what sort of toxicology information must be captured with regard to the microarray datasets in context.

DR. QUACKENBUSH: I am still learning a lot about what toxicologists do and what they think is important.

[Laughter]

So, for me, this has been a bit of a challenge but in terms of actually interpreting the data, I think what you collect has to reflect the questions that you are asking. My understanding of the toxicology field has to do with trying to predict what the response of the organism is going to be to a particular compound. So, in my view some of the things that are clearly important for understanding this are the compound, its structure because ultimately down the road we want to do data mining and what I would like to do is be able to go back and say, okay, I see this response. What I would like to do is know what causes that response. Is it compounds that interfere, are known to

interfere with a certain pathway? Or, is it compounds which simply have the right set of aromatic rings attached as what we thought were non-functional aspects or non-functional parts of the molecule? So, the compound, its structure, the dose, the time period or the time course information, information about the animal strain, genotype if it is available. I think every piece of information that you have up front is going to be valuable at a later date for mining this data and understanding the effect.

DR. WATERS: And these need to be captured in the database.

DR. QUACKENBUSH: I think they ultimately need to be captured in the database. The other thing which is very important, which people neglect, is the need for ontologies in controlled vocabularies to define these things. One of the real problems with analyzing data even in our labs when we started doing experiments, we sort of threw things out to the anarchy of the masses and let people type in their experiments. If people type in cancer or people type in tumor, and if people misspell tumor or use the British spelling of tumor and you try to extract the data from the database

without knowing what all the variants are, you only get a partial view of what is actually represented within that database. So, having standardization even at the level of experiment description and compound description is fundamentally important for later interpreting the data.

DR. KAROL: Thank you very much. We will move on to our next speaker, Dr. Ghosh, and she will be talking to us about fluorescent machine standards and RNA reference standards.

Fluorescent Machine Standards and RNA Reference Standards (Summary of Results from the NIST Workshop)

DR. GHOSH: Thank you very much for giving me an opportunity to come over here and update the subcommittee members and all the audience members on some of the efforts that we have undertaken in conjunction with NIST and industry participation in defining standards.

[Slide]

Some of the stuff which I will actually be mentioning has already been alluded to in terms of lack of standards in the gene expression area. That really prompted some of the key industry leaders, some of the NIST and FDA members, back in

2002, to get together in one of the meetings, and I will be basically outlining what was outlined for the group to achieve and accomplish.

In the second part I will cover a little bit all the activities regarding the development of the microarray fluorescent standard efforts and the working group which has now been made up of all the industry participants in terms of the fluorescent standard initiative in trying to define the specification of the standards.

The third part, of course, as we already heard is in terms of the RNA standards initiative and that group again assembled together. This was an industry, government and several academic institutions who have joined together to define what that standard is, and how it would be developed, and how it can help us to answer some of the variabilities that we are seeing today.

Lastly, some of the feedback that I got from NIST and I wanted to bring it to the table today because there is definitely a request for an active participation of FDA, requested by NIST, to really help this community and this technology to build some of these standards, and how FDA can really make an effort and contribution in bringing

that to fruition. So, I am going to present that request formally in front of everybody.

[Slide]

The kickoff meeting actually started in 2002. Fortunately, we had Frank Sistare representing the FDA over there, where we had defined that we should really look into two major areas, one being first in the scanner area which really also contributes but it was one of the easiest, less challenging perception-wise which people thought that we could actually accomplish. To be honest, we have made some very good progress in defining some of the standard needs there which I can overview for the committee members here.

So, in terms of that particular first initiative, the team got together at NIST on December 10th and, in fact, basically presented various practices which the microarray readers can adapt and define a standard and since then every month this particular working group is meeting and making progress. So, I will overview some of the definitions and specifications that have been laid down, which NIST has now taken together and they are really making that particular artifact for the community which will be available for individuals

as a calibration standard for the scanner area.

The universal RNA standard, which was the second objective laid out for the team--a meeting was held at Stanford, in March this year, and it is actually drafting a guidance document which will be out for all the participants to comment on by end of June.

The third workshop, again, was held with NIST and industry leaders in respect to the microarray fluorescent standard to accomplish the second phase of development of the scanner initiative. So, I will overview a little bit of some of the final status on those.

[Slide]

In terms of the accomplishment for the first group on developing an artifact, specifications have been developed. Currently, we are trying to define a technology which can actually accomplish the specifications which have been laid out by the working team. It is a little bit challenging because some of the finer specifications are really becoming a challenge for us to accomplish because of the dyes that we have defined and they have a finite life period. If a standard cannot be made in a way that it can be

stable over a period, it really doesn't help us. So, we are right now at the stage of defining a technology which can really give us that stability factor in the calibration standard. It is a challenge but we are right now at that particular stage.

In terms of the artifact, the draft artifact is out and it has been more or less, about 95 percent, developed but the challenge comes on if we cannot define a technology to make and accomplish those, we have to go back and change some of the specifications in terms of the available technologies.

[Slide]

The decision in the case of the artifact was that for each particular dye we will have two types of artifacts in the standard manufacturing area that people can use, one addressing the uniformity and the signal-to-noise for the right features in the scanners, and the other one will be more as a limit of detection which would be basically treated by the manufacturers and adopted in terms of the specification definement.

These artifacts won't be manufactured by NIST but an outside agency will work with NIST, but

NIST will certify and endorse it at the end of the period, and that is how the whole activity has been decided and it is totally supported by NIST in that matter.

[Slide]

This is an outline of the preliminary scanner specification decisions which the working group accomplished over a period of three to four months. Artifacts will be uniformly coated. There will be at least two artifacts per dye. The decision right now is a dye which resembles Cy-3 and Cy-5, and anything which can mimic those particular two dyes will be the first. They won't be the last but as more dyes come into the picture we will be able to adapt the same principles. The same technology which has been identified during the first initiative can apply for the other initiatives too.

Some of the major issues came up, whether glass would be the choice feature in terms of accepting as a standard and at last the committee definitely decided to go with the glass. The non-flatness of the glass in a microarray experiment, it seems like that was one of the areas, we found out, really impacts your data

quality, how flat the particular glass is that you are choosing. And we came up with that they won't exceed it than this ten micron limit because that can really alter the data quality being represented at the further end.

Various scanners right in the marketplace have different issues with this particular flatness of glass. Therefore, this was an alert figure which prompted us that many of the home-brew type of glass manufacturing may not basically understand the underlying pinning of the flatness of the glass and how it impacts the scanner reading, and how it impacts the data quality, but it is an important one.

The other part came in in terms of the thickness of the glass, flatness and the thickness of the glass, and currently this particular standard which we are going to develop will really keep to a one millimeter thickness. The artifact which basically finally came would be a 1 by 3 since the major industry is facing a 1 by 3.

[Slide]

This is a picture which defines that we have defined a particular area where the Affymetrix chip--they would basically make a cut in the major

final defined artifact slide, and use that particular region to calibrate their scanner.

So, if you look at this picture, this particular artifact can be used by 10 to 12 available scanners available today in the marketplace, and they have all actively participated in finalizing this particular design which is out there. This would be treated by the scanner as the reading zone which helps them to really scan the area, and the placement of the barcodes and the placement of the backgrounds have all been agreed to by all the manufacturers of the scanner readers.

[Slide]

A second workshop by the same scanner group was held on May 14, and the issue here was what technology we have to basically adopt. The Cy-3, Cy-5 are very unstable and photo bleaching was one of the major issues that we observed that the Cy-3, Cy-5 dyes have. Therefore, we had to look into metal oxide glasses, which are less prone to photo bleaching but currently all the available technologies really do not help us to make a particular metal oxide glass artifact which could be uniformly coated or which was uniform enough to

help us to create this artifact standard.

We have engaged now Molecular Probes, Evident Technologies with Crystal Technology as well as the Quantum Dot Technology people to come together and help us in order to define a technology whereby we could basically mimic or choose two dyes that we are looking for in order to help us to build this particular artifact. There are some experiments which have been laid down with Molecular Probes. They are currently working on it so it is in a development phase but very soon, within the next two to three months, we are trying to activate that particular activity by Molecular Probes, whereby they feel there is a particular dye. It is organic in nature, but it is much more stable than our current Cy-5 dye where we are having the biggest problem issue. So, hopefully, we will be able to identify a particular technology to help us meet our specification. Evident Technology, I would say this is a great technology to consider in terms of stability for bleaching. They are the perfect technology to adopt in terms of building a particular standard. Hopefully again, one of the dyes, they have the material available so it is not a problem. With the Cy-5 we

are struggling and time would be a factor but we are very hopeful will we accomplish that target very soon.

[Slide]

As I mentioned, these are a couple of the next steps in the scanner artifact development that we have to accomplish, defining some of the protocols and how we view the data analysis is a critical factor. It is not enough just to develop an artifact. How we use it and how we interpret the data is another area. For this particular usage, what we are looking for is a second stage of a defined protocol that every individual, not just the scanner manufacturer but individuals within the lab can basically use the protocol in the same fashion; come up with a set of metrics which would be defined. Again, technology is a big issue and there is a big variation in user terminology. What is uniformity? I have heard many definitions. And, we need unification and understanding and common consensus building in agreeing to some of these terminologies and usage.

So, we are looking for NCCLS participation in this particular last phase of activity, whereby uniform protocol and terminology would be part of

the completion of the standardization. In fact, NIST has already invited ASTM to come to the table and NCCLS to come to the table. The way we might work is that this working group may define the protocol and get it in one of their sessions of NCCLS to get some approval and understanding.

[Slide]

The next particular standards meeting happened at Stanford University on March 28 and 29. Again, government, industry, manufacturers and microarray users all collected together and shared some of their concerns, major concerns in the microarray area or gene expression area and the variations each one of them are facing. I will very quickly actually glance through some of the topics since time won't permit me to go in great detail.

[Slide]

Some of the major goals of this were educational, or providing a forum for everybody to come and share their own methods and techniques in order to define the standards for the gene expression area. There were several areas where people agreed and disagreed, but we wanted for all of them to come to the table and actually table the

disagreements so that we could hear and find out where some of the commonalities have to develop.

In fact, we were looking for a guidance and how NIST could help us in this particular initiative and participate since we look towards them in terms of the standards development, and we really need their help in order to make some traceable standards, especially from a data submission point of view too.

Requirements were laid out, like, we need to define some specifications for universally applied--some RNA standards which could be used very effectively by IND and NDA filings initially and later on as the diagnostic industry really improves, it can start building some elements there that could help some of the diagnosis and prognosis assays which are currently being developed.

[Slide]

I wanted to take a moment to really go into finer details, when we talk about gene expression, what the work flow looks like and where several of the standardization initiatives really need to happen. At the universal RNA workshop we addressed maybe some of the areas but still there are some unanswered areas. Today we heard from

John what the annotation area and data format area are going to do and provide some guidance in there.

But let's start from the very beginning, where we talked about the sample preparation area and how an RNA is extracted; how it is particularly stored; what is the particular concentration of the RNA which is put on the microarray chip. What particular integrity of the complete RNA, before even it is hybridized, how does that affect. We have found that each and every element in the sample preparation area is going to affect the data quality. So, we do need some guidance in each and every area about even the sample preparation that will be important in making final conclusions or calls at the end of the period.

For the manufacturers in the array fabrication a lot of quality control issues most probably are there, but it needs to be well understood with an idea of how it is going to impact the data quality at the end when we are doing just the data analysis. As we go through this work flow process we are accumulating all the errors as we are going through.

The effect of labeling is another part, how well we have labeled? What is the optimum

percentage of labeling that is required to give the optimum output? How balanced are the channels? We already know there are environmental effects when you work with labeled samples. How are we really taking precautions? What is the time period? What is the protocol? They need some standardization in the labeling and hybridization area.

People use different protocols in the hybridization, and they do have an impact on how we get the data at the end point. So, what is the particular hybridization protocol? How stringent is it? How well will it hybridize? Those are some of the factors--what is the cross-reactivity of the probes, and how does it affect the data manipulation at the end? We need to understand those factors.

I already talked about the scanning area, and I think the movement we have started with defining the standardization effect, it would take care of most of the scanning zone which is most promising. Then, coming to the probe area and John has mentioned a lot of these areas. Sequence homology, clone specifications and the noise, and cross-reactivity are some of the other issues that need to be developed and, again, we need some

standardization to be developed and put into place in order to have more reliable data.

[Slide]

I have talked about this, generalized work flow area. In terms of this particular Stanford meeting, we addressed the two technologies, the PCR technology as well as the microarray technology, in trying to establish a standard which can really help all the technologies. This is the common, general outline of the work flow which came out in terms of discussion. As we see, there are very generic commonalities between the two and standardization needs.

[Slide]

So, session one of our universal microarray standards--actually, Frank Sistare was our session chair and he really helped us to bring an understanding from a diagnostic perspective, what some of the standardization needs are. Maria Chen, from FDA, in fact, presented some early views on what we need to accomplish if we are really looking into some IND submissions. Again, standards were something which really popped up, that we need to develop them in order to make some relevant contribution or meaningful contribution.

Carol Thompson, from the Pharmacology Department, basically, she presented her teams and one of the projects that they are going to initiate in terms of standardization with various platforms and with mixed tissue samples in order to understand the toxicology effects across standards, and what type of standardization might be helpful in terms of protocols and interpretations. Data understanding was one of the areas that she talked about.

Some of the areas in terms of bio-international standards were brought by Merck. Roland Stoughton, in fact, talked about some guidelines, again, needing to be developed in terms of how data interpretation in the diagnosis and prognosis areas are made; how we create different standards. So, a general flavor was that for each application we might need to look into different types of standardization, but universal standards at the end of the workshop basically came out by two general guidelines of having an external standard and an internal standard.

[Slide]

I wanted to bring this experimental design which was put forth by Brenda Weiss, from the

NIEHS, whereby basically they have taken about five or six different platforms which are participating in that particular consortium.

[Slide]

The data outcome basically comes from the array platform and different labs and array to array variability trends form the maximum in terms of data variation. So, these results, which were shared, really made it very clear that unless we address the standardization needs very soon and early on with some really good participation from every segment, we will still be struggling to make some meaning out of this particular technology.

[Slide]

This is the one which was presented by Carol Thompson, from FDA, where standards for toxicogenomic studies basically would be using bench mark genes within the mixed tissue samples. Currently, that activity has already started and Frank has been actively engaging various industry participants, as well as academic participants, to really contribute to this particular project. Hopefully, some of the expected initial outcomes of this particular activity would be to identify some of the probes that can perform similarly across the

platforms. Unless we do that activity, building any databases with only one type of data may not be sufficient. It would be incomplete.

Determining the normal range of false positive and negative would be another objective of this, and lab to lab variance. Again, without some universal standards being developed, we will see a lot of variation, as being observed already by the NIEHS consortium, reported by Brenda. Ultimately, hopefully, this particular publication will be available with the findings which will help all of us to understand where we have to focus our energy.

[Slide]

The second session during our RNA development session was basically targeted towards defining some of the metrics that each of the microarray platform users needs to acquaint themselves with. These may not be just platform specific. We may need to define some metrics and RNA input sample which goes in an microarray. Some of those thoughts were basically--

[Slide]

--this particular slide shows that even procurement of RNA, when we are getting it from different sources, has impacted the data quality.

So, procurement, the source of a participant RNA, the tissue samples, isolation methods, temperature, storage, all have contributed to data quality at the end. This was a great slide, presented by Ambion, known experts in RNA. They spent a fair amount of time in digging deeper into the issues of RNA and how they have basically contributed. So, I think the metric definition part, which we have already laid out from a platform perspective, was good enough but now we feel that that is just not enough. We now have to extend it into defining some metrics, even RNA quality which is right at the beginning, and we are seeing some results coming out on how they have been impacting the data results at the back end. So, unless we define some good controls and some good specifications right at the beginning for a particular platform to address, we may not be able to interpret our data very meaningfully at the end of the experiment.

[Slide]

Going back, some of the teams from the universal RNA workshop came out with multiple sources of data variability from different technologies, from different probes and primers used by different platforms, different

laboratories, sample types and extraction methods. And, we heard it coming from every angle, wherever we looked into.

There was a great difficulty of sharing data between the platforms, and we have heard that today also. MIAME is a definite, very good start and it is being extended to the tox area. But we need to do more about the annotation problems. Unless we address the annotation issues through some work groups and common understanding, we will still be struggling to make some valuable, meaningful data interpretation.

Standards and methods for labs, which was actually very well presented, why GLP practices have always been treated as one of the areas of keen interest, we need to look into those and how each of the labs were producing these data; how they are standardizing their activities around different metrics; and how we refine our methods. That is another area I think we need to start looking into more to define and bring some consistency in our data interpretation.

[Slide]

A very interesting factor came in, which was RNA quality index. That is gaining some

momentum also now. We would eventually like to define some RNA quality index as a factor which would be treated as one of the standards as input quality RNA factor. If we have to define some of the metrics, maybe these are some of the proposed metrics which are being considered that can really make--that the metrics, when we need to define an RNA standard, we define it with particular metrics and eventually these can form our data submission pipeline.

[Slide]

So, what a good standard should be--John had actually presented the slide at our universal RNA standards workshop--what it should do. It definitely should be something that could be used by a platform over time, compare between the different platforms; should be consistent enough, therefore, some of the concerns of using biological samples as a universal standard were basically thought through and we couldn't find the number three parameter, that it has to be consistent over time. We thought that most probably we might have to go to synthetic model having all the biological characteristics for that standard so that consistency can be maintained over time. We should

have a well-defined protocol. That was definitely one of the themes that ran across and people agreed that a defined protocol needs to come out through that activity. And, we must be able to make both absolute and relative measurements using this particular standard. It should not just be confined to use in the gene expression but QRT-PCR technology should be able to use that.

[Slide]

What are some of the microarray performance characteristics? From a design and fabrication point of view, platform types. The surface types which are used by fabrication and a manufacturer may impact in terms of data quality; understanding each and every aspect of the surface types. Composition and spatial layouts, a number of replicates identifying that particular array can be some of the very good requirements that can be laid out during submission of data. In terms of the spot elements on a microarray, clones, sequence, primers, probe lengths, gene name, etc., can basically be added to the list of spot element definition. Built-in controls, which are the housekeeping genes for the controls defined by an array manufacturer, can be defined in terms of

requirements.

Again, in the microarray controls area, use of internal controls, which can be synthetic housekeeping genes; pooled RNA from sample cell lines or pooled RNA from test samples; and RNA and oligonucleotides from plants and bacteria can also form microarray controls. But these were some of the controls that we saw came out of the meeting that individuals presented.

So, there is a lot of different variation where people have been working. Because availability of a standard is missing, people have been trying to use some of the internal controls but it seems like it comes that we do now have to come up with a unified defined protocol for all this.

So, standards are required for several purposes. This was the proposed workshop recommendation, that periodic laboratory proficiency testing can be used for platform performance validation and baseline monitoring; cross-platform performance validations and inter-laboratory performance validation. These are some of the themes that would be basically addressed as we define the external standard

through this work group.

A consistent definition of terminology, which was pretty varied, and through the guidance document this particular definition of terminology part would be addressed so we can define a consensus for how we can define the terminology. Finally, the consensus of the attendees at the end of the session was that there has to be an external synthetic RNA standard reference and an internal RNA standard reference which would be treated as a spiking control.

[Slide]

These were the two particular standards which were defined by the work group. The definitions and the specifications of the RNA standards are coming out, as I said, in a guidance document which will help us. In terms of the reference method, we most probably again have to engage external agencies, like NCCLS and ASTM, to work with NIST in order to define the reference standard method.

[Slide]

I want to go to my last slide. Here are some of the open questions which came up at the end of the session. NIST had taken up this particular

initiative to define the specification for the work group but the next phase of execution and implementation plan, they are really requesting FDA to come to the table and define their requirements, and they are proposing a partnership model with the industry to take place in order to execute it. So, I wanted to formally place that requirement, as per my discussion with NIST on Friday where they made this requirement. They are ready to come and sit with FDA and take the requirements from FDA so that they can work to a particular objective which will help FDA to accept the data. That would be the next step. Frank has really been helping this particular activity and bringing all the feedback to the table to help really guide us on what should be our next step and how we should address that.

With that, I will address any questions if the committee has any questions.

DR. KAROL: We will just take one question.

DR. ZACHAREWSKI: In the open questions you said that the guidance document was going to be published by the end of June, 2003. That is in a couple of weeks. Is that still on schedule?

DR. GHOSH: Right, it is on schedule. It

is written up. It is waiting to go to the session chairs, and John Quackenbush was one of our session chairs and Frank was one of the session chairs. We have two other session chairs who need to review the document and give their comments in terms of completion.

DR. ZACHAREWSKI: And where will that be published?

DR. GHOSH: It will be published by NIST actually.

DR. ZACHAREWSKI: How will it be available?

DR. GHOSH: All the activities of the standards workshop are currently available on the NIST web site. So, this particular guidance document will eventually go up on the NIST web site.

DR. KAROL: Thank you very much. We appreciate your presentation. In order to be able to fit adequate discussion and the open public hearing, we are going to change our agenda just a bit. We are going to break for lunch now and reconvene at one o'clock after lunch.

[Whereupon, at 12:15 p.m., the proceedings were recessed until 1:00 p.m.]

A F T E R N O O N P R O C E E D I N G S

DR. KAROL: I would like to start the afternoon session. First is the open public hearing but there is no one scheduled to speak so let's move on to Dr. Leighton, who is going to talk about the CDER IND/NDA reviews.

Topic #3 CDER FDA Product Review and Linking Toxicogenomics Data with Toxicology Outcome CDER IND/NDA Reviews - Guidance, the Common Technical Document and Good Review Practice

DR. LEIGHTON: Good afternoon.

[Slide]

I will spend the next few minutes providing a general overview of the CDER IND/NDA review process and describe the nonclinical studies that are usually submitted to support these applications. I will also spend some time discussing the role of FDA and INCH guidance in the review process; a slide on the common technical document, as well as the CDER pharmacology good review practice.

The purpose of my presentation is to present to you the current review practice and to introduce a possible future role of pharmacogenomics in safety assessment, and this is

not intended to be a complete discussion of the review process.

[Slide]

The review team for any IND and NDA consists of the professionals shown on this slide. It includes project managers that are the first, and sometimes the only contact that a sponsor has with the division; medical officers; pharmacologists, toxicologists; chemists that examine the manufacturing process; and clinical pharmacokineticists and statisticians. Now, the first four disciplines are primarily involved in the initial IND review. Clinical pharmacokineticists and statisticians are brought into the review process on an ongoing basis as needed.

[Slide]

The nonclinical studies usually submitted to support an IND and NDA are shown on this slide, including studies on the mechanism of action, such as pharmacodynamics and pharmacology studies; studies on pharmacokinetics, including absorption, distribution, metabolism and excretion; safety pharmacology studies which are studies that provide an evaluation of vital organ function, in specific,

cardiovascular, central nervous system and respiratory function; general toxicology studies that provide the pivotal safety data for an initial IND. Genetic toxicity, reproductive toxicity and carcinogenicity studies are also provided.

[Slide]

The goals of nonclinical IND studies are primarily at the initial stages, number one, to identify an appropriate start dose; secondly, to identify organ toxicities and their reversibility; and third, to guide dosing regimens and escalation schemes.

[Slide]

Pharmacology studies--pharmacologic activity as determined by in vitro and in vivo animal models, and nonclinical studies are generally considered of low relevance to the current safety assessment as provided in the IND and efficacy studies in the NDA, which is primarily determined by Phase III clinical data. Therefore, for this reason, summary reports, without individual animal records or individual study results, usually suffice for reporting requirements for pharmacology studies.

[Slide]

However, toxicology studies provide the pivotal information for the initial safety assessments, as well as the start dose decision. Ideally, toxicology studies should mimic the schedule, duration, formulation and route as that proposed for the clinical trial. They should conform to standard toxicology protocols and should be conducting according to good laboratory practices, or GLPs, as identified by Code of Federal Regulations, Section 21, Part 58, or 21 CFR, Part 58.

[Slide]

To support an initial IND what should be provided? An integrated summary of the pharmacology/toxicology data should be provided. Unlike that I described earlier for pharmacology data, a full tabulation for each toxicology study, including individual animal data, should be provided to the review divisions in order to support the safety of a proposed clinical trial.

How can pharmacogenomic data be incorporated into the initial IND safety assessment? Well, perhaps this data can be used to assist in the selection of a start dose, a choice of a relevant species for additional long-term

studies, or to identify biomarkers for future clinical evaluation.

[Slide]

Not all toxicology studies need to be provided with the initial IND. It is an ongoing process that should be conducted concurrently with clinical develop. So, some of the studies that may be provided, and this depends to some extent upon the intended indication for the drug--some of the studies that could be provided at a later date include long-term toxicity studies. The genetic toxicology panel should be completed if it hasn't been completed by the initial IND. Reproductive toxicology studies should be provided, and carcinogenicity studies should be provided if the indication and the treatment warrants them.

So, how can pharmacogenomic data assist at this stage? Possibly by decreasing the study length. For example, carcinogenicity study standard is usually a two-year rodent bioassay. Perhaps now, with additional pharmacogenomic data, studies can be conducted in a shorter duration, perhaps six months. Improve assessment of organ toxicity in terms of clinical relevance, and provide mechanistic explanation of toxicity.

I would like to emphasize that at least initially it is unlikely that pharmacogenomic data will replace the standard assessment. For example, in general toxicity studies there is usually provided histopathological evaluation of over 50 tissues. Most pharmacogenomic studies only look at one, two or maybe even a handful of tissues. So, it is unlikely that the data will be of sufficient extent to supplant our traditional general tox environment.

In addition, one other point is that the animals often die in the middle of the night. It is very inconvenient and you may get a lot of tissue autolysis and with the issue of RNA standards being critical, how will this RNA look in the morning when the animals are finally found and the tissue is extracted? So, the cause of death may not be amenable to understanding by genomic analysis.

[Slide]

What is the role of FDA guidance in the review process? ICH stands for International Conference of Harmonization. FDA/ICH guidances represent the current thinking of the agency. These are recommendations, not requirements. And

FDA guidance can either be drafts, which is for comment purposes only, or final documents. So, it is a step-wise process where the agency can get the input of outside experts. Guidances are available on the CDER web site.

[Slide]

Some of the FDA/ICH guidances, on the left-hand side are process-driven guidances. These include things like guidances on how to submit an IND; how to select an appropriate start dose; how to design an appropriate study for acute toxicity testing; and how to submit an electronic NDA. On the right-hand side are some guidances, and this is not a complete list but some of these guidances that are available include some more scientific-based guidances, including guidances on carcinogenicity dose selection; genetic toxicity; reproductive toxicity; photo safety testing; immunotox; and biotechnology.

[Slide]

One of the guidance documents that are available is the common technical document. This is a guidance that describes a harmonized format for technical documentation for registration in all three regions. By the three regions I mean United

States, the European Union and Japan. This is for registration so this would be for the NDA stage. It consists of five modules. Modules two through five are common to all regions. Module one would be region specific. The purpose of the common technical document is to reduce the time and the resources used to compile a registration document. It is intended to be used with other ICH and agency guidances and to allow for regional specific summaries.

[Slide]

In an effort for transparency, the pharmacologists have developed what is called the good review practice. This is a guidance for reviewers and provides for a standard review format. It is an internal review format for the IND and NDA primary pharmacology reviews.

The purpose of this good review practice is to provide for standardization of reviews across divisions to ensure that important information is capture in all reviews, and it allows for continued assessment of an IND. It is consistent with the common technical document that is available at the wed site at the bottom of the page.

[Slide]

Some of the information that is collected in a good review practice, currently collected as part of a general toxicology study review, includes the information shown on this slide. It evaluates mortality, clinical signs, body weight, food consumption, ophthalmoscopy, electrocardiography, hematology, clinical chemistry, urinalysis parameters, organ weights, gross pathology, histopathology and toxicokinetics when they are available.

[Slide]

In summary, there is a different submission format provided for pivotal safety data, in other words your toxicology data, relative to pharmacology data. We have developed good review practices for the evaluation and capture of data in order to provide consistency among review divisions and to increase transparency. Good review practices, if they are developed for pharmacogenomic data, will need to consider the interdisciplinary review of pharmacogenomic data that was discussed earlier by Dr. Woodcock. It is my belief that pharmacogenomic data will play an important role in the safety assessment in future INDs and NDAs. Thank you.

DR. KAROL: Thank you very much. We will have questions at the end of this session, after the four speakers, so we will move right on to the second speaker. This is Dr. Levin who will talk about electronic submissions guidance, CDISC and HL-7.

Electronic Submissions Guidance, CDISC and HL-7

DR. LEVIN: I am going to be talking about some of our standards development and implementation at FDA.

[Slide]

I am going to go over some of the standards organizations that we work with at the FDA, the FDA Data Council inside the FDA but then there are these four other organizations I will be covering. I would like you just to concentrate on these four organizations, right here, and see if you can find a pattern in all those initials and see what the next organization should be after this one.

[Laughter]

I will go through what all those abbreviations stand for. I have three initiatives here but I understand we are a little pressed for time so I am going to go over two initiatives, the

clinical and nonclinical study data standards and the annotated ECG waveform data standard. I will describe why those things are important here.

[Slide]

We deal with a number of different standards development organizations inside the government, accredited standards development organizations and a variety of other standards organizations that are not accredited.

Inside the government we have the FDA Data Council. We also work with a group called Consolidated Health Informatics. For accredited standards development organizations we work with Health Level 7, which is accredited by the American National Standards Institute, and then two other standards groups that we are working on with ICH.

[Slide]

The FDA Data Council is what we have formed inside the FDA to try to standardize across our various centers. We have the Center for Foods, Drugs, Devices, Biologics and Veterinary Medicine so we try to standardize across these different groups to have standards that are common in the FDA. We have representatives from all the various centers as well as the different offices and the

Office of the Commissioner. This group is involved with the national and international standards development.

[Slide]

Here, in this group, we coordinate the standards development. We get information that is coming from different centers or offices where they want to have data or terminology standards. We form expert working groups within the FDA, work on the standards, work with standards development organizations if there are already standards created or, if we create our own standards we try to bring them to a standards development organization, like HL-7.

[Slide]

There is another group we work with, the Consolidated Health Informatics. This is a group that is part of the President's eGov initiatives and it is to set the standards for inter-agency use. There are three major partners in this organization, Department of HHS, Department of Defense and the VA. So, those are our three major partners in this and what they are trying to do is set standards that can be used across the different agencies in health care. This was started because

the Department of Defense and VA were trying to exchange information and were unable to because they use different terminology and they said we are going to use the same terminology and form this group. All the government agencies that deal with health care are involved with this group.

They have set five standards so far. One is to use HL-7, Health Level 7, for messaging standards. The other is to use logical observations, identifiers, names and codes, LOINC, for lab test standards, and use DICOM for transmission of images, and the National Council of Prescription Drug Products for prescription messages and IEEE for ECG monitoring messages. So, these are some of the standards that they have. These are the first five. They have now listed 24 different standards groups that they want to establish and they are moving forward on that. Once these standards are established, that means these government agencies will use these standards for exchange of information. The first two are important to the FDA, the other three are more related to agencies involved directly with health care but there are other standards that will be coming forward that will be important for us when

we are dealing with research and the other things that we deal with as we interact with drug companies and investigators.

[Slide]

Health Level 7 is an ANSI accredited standards development organization. They are an international group. They have open membership. They follow all the procedures laid out by ANSI so that their standards are accredited and they can be accredited by ANSI or ISO. They are involved with standards development activities in the government. They were involved with the Health Insurance Portability and Accountability Act which provides standards for exchange of insurance information and prescription drug information. They are involved with the national health information infrastructure which is to develop standards so health care groups can communicate information. They are labeled as the standard message for the Consolidated Health Informatics group.

FDA is part of the Health Level 7. We are on the clinical research information management technical committee in Health Level 7, and this is where standards that are of interest to the FDA would go for accreditation. So, we take our

standards to the HL-7 group and we have taken a number of standards there for development and subsequent ANSI accreditation. We are also involved with the vocabulary technical committee where terminology standards are being looked at. Since there is a lot of government involvement in Health Level 7. We are involved in the government special interest group which includes groups like the Department of Defense, VA, CDC and NIH.

[Slide]

John was just talking about ICH. We are involved with that. There is the common technical document, as he was describing, as well as some terminology through ICH. There is something called MedDRA, which is terminology for describing adverse events, and we are using that for exchange of individual case safety report information.

[Slide]

Finally, there is a group called CDISC, the Clinical Data Interchange Standard Consortium. This group is an open group. Though they are not accredited, they joined HL-7 so they are involved with HL-7 as well. There are representatives in this group from vendors, pharmaceutical companies, industry consultants and government agencies. They

are trying to develop standards for clinical trial data between pharmaceutical partners and between the pharmaceutical companies and regulatory authorities. They have set forth a standard, what they call a submission data model for submitting clinical data, research data to the FDA.

[Slide]

These are the standard initiatives that we have brought forward, that we are working on right now. There is one for electronic submissions of applications; study reports; structured protocols; a standard for product labeling; a standard for individual case safety reports; electronic MedWatch; stability data; annotated ECG waveform data; and study data.

[Slide]

Now I will just briefly go over two of our standards. One is the one for clinical and animal study data. The clinical study data comes from the CDISC group. The animal study data we are working on is a separate group but it was facilitated by the CDISC group and this has been following the same basic standard that was worked out with the clinical standard, which I will go over.

What I am going to talk about is a

standard that is based on the CDISC version three, and this is available on their web site as CDISC.org if you want to find out more information about that. The standard development is divided into two parts. One is the submission data model and the second part is terminology. What I am going to describe now is just the part we are working on now, the data model, not the terminology which we haven't really gotten into. What we are working on also is standardization procedures, including the development of specific analysis tools and a data repository for this type of data.

[Slide]

The CDISC version three data model divides a study into a collection of observations, and there are three types of observations, interventions which are therapeutic or experimental treatments; events, which are incidences that are independent of the planned study observations, for example adverse reactions; and findings, which are observations resulting from planned evaluations to address specific questions.

[Slide]

Each observation is characterized by a set of descriptive variables. There is a topic

variable which identifies the focus of the observation. There are identifiers which identify the subject or the study uniquely. There are timing variables that describe the start and end of an observation. There are qualifiers that describe the trait of an observation.

[Slide]

Here is an example of an observation in a clinical trial. This would be the topic of the observation. The identifier, subject 101A is the identifier. Starting on study day six would be an example of the timing variable, and that it was mild would be an example of the qualifier. There is a series of these variables to describe the different topics, identifiers, timing variables and qualifiers. So, this is what the model consists of, a series of these descriptive variables to describe observations.

[Slide]

The other standard that we are working on that might be relevant to this discussion is the annotated ECG waveform data standard. This standard is also brought through HL-7 and is based on their reference information model, and is an XML file.

The interesting part about this data is it represents the digital ECG with all the annotations that the company would put on the ECG--where the P wave starts, the QT interval duration and things along those lines. But it is a large amount of data since it records every point along the line of the ECG. It really was started off as a correlated data standard or way to transport correlated clinical data or study data. So, when we looked at this model, since it is transporting a tremendous amount of information that is correlated, this might be something that might be useful for the data that we are discussing here.

This data, along with the clinical data model, are two things that we would have to coordinate as we are working with our data standards so that whatever way we decide on transporting this information is related to a standard that is coordinated with everything else that we are doing, and we would like to take it through the different standards groups so that we are coordinated with the other parts of the research community. Thank you.

DR. KAROL: Thank you very much. We will move right on to Dr. Mattes, who will tell us about

MIAME-Tox.

MIAME-Tox

DR. MATTES: In truth, I am going to be talking about MIAME-Tox in context of a larger issue, much of which has been covered before and I am probably going to rehash quite a bit but I will try and make that fast.

[Slide]

The larger issue is that of the ILSI-EBI collaboration which has been a learning experience for both of us in terms of handling toxicogenomic data.

[Slide]

Again, I am going to kind of come at a pretty high level and talk about why we need a database, why it is essential; how we envision that it is going to be developed; what are the issues; and who is involved, particularly the ILSI-EBI collaboration.

[Slide]

Just to reiterate kind of one of the issues which I think is the most significant issue, and the most significant issue is how we were trained X number of years ago, even maybe five, ten years ago, to think about biology. In fact, we

were trained as graduate students and post docs to look at one tree at a time, focus down and analyze it and write up your thesis along those lines.

[Slide]

"Omic" biology--genomics, proteomics, whatever, really, unfortunately or fortunately, or whatever, the characteristic is looking at the forest and mountains, the big landscapes and trying to discern from that what is going on. Yes, things do happen in individual trees but the data can't be addressed at that level. So, the way forward is really with informatics. Quite frankly, I think it forms a stumbling block for most people and it is very hard to fully integrate your thinking along the lines of informatics as the way forward.

[Slide]

Again to reiterate why you need to handle this sort of data in a database, if you think about the traditional endpoints that are accumulated per animal it is, you know, dozens, whereas genomic endpoints in any given animal is going to be thousands.

[Slide]

But there are other issues, and there are other significant issues that can only be addressed

at an informatics level. One is the influence of the technology. I have spent a fair amount of my time getting hung up on the informatics of sequence analysis and I am passionate about that because it really influences the endpoints, the measures you are getting.

[Slide]

I give as an example that many genes are alternatively spiced and these events are not usually unambiguously detected by microarray.

[Slide]

I give as an example a classic one, which gives the all too famous UGT1 gene which consists, when it is spliced, of five axons that are spliced together but there are six alternative axons which result in six different proteins from this one gene, if you will. Yet, when you think of array technology most arrays are going to be targeting the 3-prime UTR that is just sort of technologically driven. So, all too commonly you may think you are measuring one sequence but, in fact, you may be measuring something else.

[Slide]

On another level, for most cDNA arrays you have to address the issue of whether or not the

probe may hybridize to more than one sequence, and the bottom line is that you have to have a database that captures the probe sequence to resolve the discrepancies between array platforms at the level of sequence. There is just no way it is going to be done manually.

[Slide]

How are we going to develop the databases? The efforts that have already been put forward were organized by what is called the Microarray Gene Expression Data Society, or MGED. They have come up with a number of key concepts. The first is this MIAME, the minimum information about a microarray experiment. I have quoted from the MGED web site how they describe that but it is essentially what should go into the database; what is the minimum information u need to be able to make sense out of the results.

[Slide]

The basic areas that are covered in this are the experimental design, samples used, the extract preparation, labeling, the hybridization procedures and parameters, measurement data and specs and the array design. Now, truth be told, all of this is focused around the original MGED and

MIAME focus which was not toxicology. It was more looking at array experiments that would come with kind of a minimal amount of biological descriptors.

[Slide]

The MGED Society also came up with MAGE, and I should say MAGE-ML. Under MAGE there is more than just MAGE-ML. These are the programming conventions and the data structures to be able to communicate the data. So, you have a MAGE-OM, the object model for the data. Then you have a markup language which allows the exchange of the data from one database to another. So, really what MAGE is about is structuring your data and structuring a way to communicate your data such that, quite frankly, as long as you have a MIAME compliant database it doesn't matter whether or not you use your database or somebody else's database, the data should be able to transfer seamlessly back and forth.

[Slide]

Finally, under the MGED Society--not finally, there is another point but under the MGED Society is an ontology working group which is striving to provide a vocabulary that will communicate the information about a particular

topic, in this case microarrays, but it is also not just communicating the knowledge but allowing its interpretation and use by computers. That is an important point because when we say, in the example that was given earlier using two different spellings for tumor, the British and the American, anyone in the room would understand what that is but, one, if the computer wasn't trained to recognize the synonyms or there was only one way forward on that, one of those would cause serious problems. So, it is not just communication from person to person; it is communication from computer to computer in a way that the computer can make sense out of it. So, if you do have an ontology that has standard terms, what you allow are structured queries and unambiguous descriptions of experiments.

[Slide]

John Quackenbush is a representative from this angle of the MGED Society. There is a data transformation and normalization working group which is striving to establish standards for recording how the microarray data is transformed and normalized.

[Slide]

So, what about toxicogenomic databases? What are the issues here? Well, first I want to throw out an overview where the ILSI effort is. Again, you have probably heard some of this but just as a recap, in the genotoxicity group there are upwards of 10 array platforms, 11 compounds with two time points and up to 10 doses per compound--it is fair to say, a fair number of arrays. Nephrotoxicity group, six array platforms, three compounds, a total of 260 animals. Suffice it to say that 260 animals means that there are at least that number of array data points in there.

[Slide]

In the hepatotoxicity group they used about eight platforms, two compounds, a total of 144 animals. In this case, those 144 were split into two in-life studies per compound. Now, for all of the groups there was analysis of each sample at multiple sites. So, the ILSI effort really represented I think a microcosm of the kinds of issues that are going to be confronted when folks try to pool data together from multiple sources.

[Slide]

One of the issues going into this we really fully unappreciated was that MAGE, MIAME or

MGED ontologies just did not address the traditional toxicology endpoints, the issue of organ weights, clinical pathology, histopathology and the like. That was not specified in the original MIAME document or the MAGE-ML. So, that became an issue for ILSI and EBI to address.

[Slide]

Likewise, another issue is that these tox endpoints are standardized in nomenclature. We have heard that referred to before. I have dug up at least two types of nomenclature for clinical pathology and chemistry. Under histopathology, this is at least the length of the list and who knows there are groups using their own customized list as well. For putting together the ILSI-EBI database we chose to work with the IUPAC designation for clinical pathology and we borrowed, if not stole, liberally from the NTPs TDMS pathology code database.

[Slide]

I keep referring to the ILSI-EBI effort but I think it is important to remember that it is not occurring in a vacuum, nor is there a lack of other players out there. A number of private companies have put together toxicogenomic databases

with a variety of different foci. Genelogic, Iconix and Curagen are the main players in this. Tim Zacharewski's lab at Michigan State has published a database structure that is designed to handle toxicogenomic data. It is called dbZach. Mike Waters' group at the NIEHS is putting together a database referred to as CEBS, which is Chemical Effects in Biological Systems. NCTR has also developed a structure to capture array data, called ArrayTrack, and last on the list is the effort that ILSI partnered with EBI.

[Slide]

The collaboration came out of one of ILSI-HESI's goals as far as the genomics subcommittee. That was the establishment of a database for toxicogenomics data. Indeed, these three bullet points are the ones that we were charged, in the database working group, to push forward on. Importantly, and I think this is an important point, we wanted the database to be able to interrogate the gene array data and integrate it with genomic experimental and toxicological domains. That would gain knowledge of links between gene experiments changes and toxicological endpoints. This is a key point because I would

venture to say that while you have heard discussions and often hear discussions of people looking at array data and saying I see a correlate with a biological endpoint, usually that correlation is made, quite frankly, sort of by human intuition, in other words, at the high dose group I saw certain histopathological effect and I see the gene changes so, therefore, there is a correlate. Or, let's say a particular group had on the whole an elevated ALT level and that correlated with on the whole the gene changes we saw for that group.

What we are trying to drive to here is to be able to do that kind of correlation on a statistical, electronic and individual animal basis within the database. So, the thrust of it and the challenge is a little bit beyond that essentially intuitive approach to those correlations. It is an approach that would get you to answering certain questions. I will get to that in just a minute because I just want to mention some of the issues that we have in the collaboration.

[Slide]

We needed to provide a way to integrate the different domains. We needed to control the

annotation. Of course, you need to centralize the information. You need to improve the array annotations as genome assemblies are released and improved, and allow data comparison. That gets to the point that you want to be able to go and compare data from different domains.

[Slide]

I think my point here is just simply that we needed to get internally consistent data to be able to run these complex queries and, yet, we had data emanating from several different sites.

[Slide]

Here is the meat of the question, a simple question, does gene X expression go up after treatment with compound Y with biological endpoint Z in experiments from ILSI members A and B? That is relatively easy to ask. You look at gene X, you look at biological endpoint Z and, you look at compound Y, and you look at a couple of datasets.

However, it is not a simple question. One that you can only address with the databases, is one which follows: Which are the most reproducible gene expression changes for all the experiments on the array with biological endpoint X, and which functional category do these genes belong to and

which are the human homologues? That is a challenge and it simply requires you to have a robust database where the data is captured in a standardization way and mapped on the sequence level.

[Slide]

Which brings me, since I am talking about standardization, to MIAME-Tox. MIAME-Tox is simply an international effort to share expertise, encourage harmonization and promote a standardization initiative. So, with the central theme being toxicogenomics, this represents an alliance between ILSI-HESI, EMBL-EBI and, quite frankly, Mike Waters' group at the NIEHS, at the National center for Toxicogenomics. It has been an extremely fruitful effort so far and I would say that this is a party that is growing and we are encouraging folks to join in.

[Slide]

These are the objectives. The first is to come up with standard contextual information. That is, put together a worldwide scientific consensus on what is the minimal information or descriptors you need for array-based toxicogenomics experiments.

Another objective is that of data harmonization, how you encourage use of controlled vocabularies for the toxicological assessments. Another objective is to push for data integration and data sharing so that you can link data within a study or several studies from an institution and exchange datasets among institutions. Finally, to set up a structure for data storage that will allow the development of data management software and databases. Right now, the two that we are talking about in development are ArrayExpress at the EBI and CEBS at the NIEHS National Center for Toxicogenomics.

[Slide]

There is a document out there to promote standard contextual information. It is trying to define the core common to most experiments. It is designed to promote data harmonization, capture and communication. Along those lines, in terms of this harmonization and communication, it is worth remembering that MIAME-Tox is based upon the same structure that MIAME has. However, MIAME-Tox document really is a focus on the toxicological domain, the sample treatment and conventional toxicology information as it is integrated with the

microarray information.

[Slide]

You can look at this document at either the MGED Society web site of the ILSI-HESI web site, and it is really out there for circulation, for review and for comments. The MIAME-Tox group is working closely with the MGED working groups, in particular the ontology working group, with the thrust of trying to develop controlled vocabularies.

[Slide]

In our hands, really what we were confronted with for this controlling data and controlling the structure and nomenclature was to look at data input as a key step. So, with the charge of capturing data in a standard manner, EBI developed what they call the Tox-MIAMExpress. This is used to store information domains in a database, the ArrayExpress database, and allow comparing queries across and within domains.

[Slide]

I am going to kind of quickly go through some Tox-MIAMExpress web shots because I think to take a look at this gives you some sense of how the data is organized, how it is going in. First you

have a protocol submission which really covers not just the microarray experiments but, obviously in the case of toxicology now the conventional toxicology tests. So, you can see here are the kinds of protocols that you can submit. Obviously, once you submit one you can refer to it for any experiments that use that protocol.

Then you move on to the array design submission which is important because these are the procedures that format the array design into something that EBI database can use to refer from one array to another. It also sets up a set of procedures to re-annotate or update your array designs via link to sequence data at EBI.

The experiment submission is now actually the meat and potatoes of it where, first, you are going to submit the experimental design, some of the information about quality controls and, finally, the samples. Quite frankly, the samples are your individual animals.

The point that follows is to submit toxicological endpoints, what sort of extracts you make from individual tissues, what sort of labeled extracts are going to be used for microarray data and finally the hybridizations that are used for

the microarray data.

[Slide]

This gives you a screen shot of the data that we have been entering into it. Obviously, you can get a flavor for what kind of data is captured, how it is captured. The drop-down menus allow control of the vocabulary. I venture to say, after working through this personally, it is a work in progress. It captures a great deal and represents I think a fantastic starting point but it is something that I encourage everyone in the audience, and anyone out there, to offer input on.

[Slide]

Here is an example of data entry for clinical pathology. The challenge, of course, as we have found in our own hands, is if you have collected the data in different units and you have to convert them.

[Slide]

These are the sorts of clinical observations that are collected.

[Slide]

I would like to add something to this slide, and that is some of the future directions but first I want to say where we are with the

status. I have shown you the interface and the infrastructure that is already in place. I have alluded to the fact that it is not as if it is fixed or immutable at this point. We are putting data into it. It is not complete yet but we envision that probably in the next quarter or so.

There are some key important points I want to mention in terms of future development. Certainly what I have alluded to is developing the tools that will query across different domains. That is not listed in this slide but it is definitely something that we are looking to work with EBI on. Finally, a key point in further development is working towards automated data upload or electronic data upload of toxicological data. That is, if it is already collected in an in-house electronic database, how can we transfer that data seamlessly using an electronic upload?

[Slide]

I would like to end with some mention of the guilty parties. Certainly, the Microarray Informatics team at EBI and Alvis Brazma is the MGED Society president and really I would say one of the MIAME proponents. Susanna Sansone has been our key contact at EBI and responsible for really

all the progress you have seen in the database there, with Philippe Rocca-Serra helping her in putting that together. I don't have Mike Waters' name here but I should because he has been an invaluable help in contact at the NIEHS. Of course, the rest of the EBI steering committee has been an important player and, finally, certainly the genomics committee. With that, I thank you and will take questions.

DR. KAROL: We will take questions right after the next speaker. So, our last speaker in this session is Lilliam Rosario, who will talk to us about CDER FDA initiatives.

CDER FDA Initiatives

DR. ROSARIO: Good afternoon.

[Slide]

My presentation today will basically address four main initiatives that CDER has undertaken so far in an attempt to better understand the field of pharmacogenomics and to anticipate regulatory considerations stemming from the rapidly evolving field of toxicogenomics.

[Slide]

So, what I would like to do is tell you about the formation of the nonclinical

pharmacogenomics subcommittee. I also would like you to know about some of the regulatory research lab-based initiatives currently going on stemming from the Office of Testing and Research. I also would like to tell you about ongoing collaborations with Iconix Pharmaceuticals, the developers of a drug matrix of microarray data linked to tox parameters and, finally, our collaboration with Expression Analysis to come up with a mock submission of microarray data provided by Schering Plough.

[Slide]

First I would like to tell you about the nonclinical pharmacogenomic subcommittee. The subcommittee is part of the pharm/tox coordinating committee and has been founded to address the rapidly developing field of pharmacogenomics. The goals of this committee are to recommend standards for the submission and review of nonclinical pharmacogenomics and toxicogenomic datasets to develop an internal consensus regarding the added value, the best interpretations in drug development and regulatory review implications of this type of nonclinical data, and to develop Center expertise and an appropriate infrastructure to support the

review of these types of data. I also should note that the objectives of this committee may continue to evolve with time to include, for example, proteomics and metabonomics.

[Slide]

The membership of this committee is intended to be very broad and currently it has participants from all the different ODEs, the Office of Testing and Research as well as the Center for Biologics.

[Slide]

The functions of the subcommittee are to interface with other CDER review disciplines, such as the clinicians and the statisticians, and other centers within the agency in recommending review standards. It is also to develop specific initiatives to keep committee members abreast of the latest developments; to assist other submissions and center groups in developing educational opts in pharmacogenomics and toxicogenomics; to provide forums for communication to regulated industry; to obtain external expertise to evaluate the scientific developments, as well as to provide internal expertise in evaluating nonclinical data submissions that contain

pharmacogenomic or toxicogenomic information.

[Slide]

This committee was formed last August and it has been extremely active since then. So far it has contributed input to CDER mg concerning research information package and no regulatory impact, as you heard from Dr. Woodcock this morning. It has contributed to the nonclinical section of the CDER draft guidance on pharmacogenomics and pharmacogenetics, and initiated process toward the development of a draft guidance on the content and format of nonclinical pharmacogenomic data submissions, and this is one of the reasons why we are gathered here today.

It is currently actively participating in collaboration with Iconix Pharmaceuticals, and I will tell you a little bit about that collaboration further on, and participates in the collaboration with Expression Analysis and Schering Plough. So, as you can see, this subcommittee has poised itself to really serve as an interface within the agency to provide internal expertise and to seek out expertise from outside collaborators.

[Slide]

I would also like to tell you about some

of the regulatory research lab-based initiatives. These are aimed at really getting the technological part of microarray data to bring it into regulatory practice.

[Slide]

It has done so by an early active participation in the ILSI collaborations, and this will be nephrotoxicity and genotoxicity working groups; collaborations with Affymetrix and Rosetta, and this will be with the cardiotoxicity focus; also collaborations with NCTR and Schering Plough.

[Slide]

As was mentioned before, these lab-based initiatives are trying to get a handle on all the technology issues. For example, genome scale expression data submitted to the agency could be generated from a variety of microarray platforms, and these platforms can be from oligonucleotide or cDNA-based arrays, numerous commercial platforms as well as in-house custom arrays. So, one of the big questions is can a standard be developed that would help assure the FDA of the biological truth, that is, the biological truth independent of a platform and site or processing?

[Slide]

As you briefly heard from Dr. Ghosh, there is an ongoing project through the FDA Office of Science and Health Coordination. It has funded a collaborative project to evaluate performance standards and statistical software for regulatory toxicogenomic studies. This study as a laboratory component that is headed by Drs. Thompson and Fuscoe from CDER and NCTR respectively. It has a laboratory component with outside collaborators that include Rosetta, Agilent, NIEHS, Amgen, Iconix and Affymetrix, and it has a statistical component that is being provided by FDA centers.

[Slide]

The goal of this project is to generate and evaluate a complex mixed tissue standard's utility for assessing platform features. What will be assessed in this case will be to assure that there are no manufacturing defects; that there is insignificant platform lot-to-lot variability; to assess the integrity of feature location; to ensure that there is unambiguous consensus sequence annotation; and a lack of cross-contamination in tiled probe features.

[Slide]

The standard will also serve to assess

experimental performance. I won't go through all these points but just tell you that these will be aimed at assuring that the biological conclusions are independent of the platform and represent the biological truth.

[Slide]

Again as Dr. Ghosh mentioned earlier, the proposed steps for testing the feasibility of a mixed tissue standard is by using bench mark genes, in this case to identify tissue-selective, low variance housekeeping genes from control animal data in large databases, and to select the tissues with most consistent expression among control animals and most coverage of the probes.

[Slide]

As you can see, we also have a laboratory component that is trying to sort out the technological issues in order to bring this new technology into regulatory practice.

[Slide]

I briefly want to tell you a little bit about our collaboration with Iconix Pharmaceuticals. Iconix Pharmaceuticals are the developers of the DrugMatrix that contains microarray data that is linked electronically to

toxicology and pharmacology endpoints. So far, Iconix has provided research access to the DrugMatrix system for evaluation purposes to train members of the subcommittee.

We visited their facility back in January and they provided some training, and continue to provide support and understanding in working with their database. They have provided us with hands-on experience using a chemogenomic data and tools, including the application of molecular toxicology markers to predict drug actions. Also, we got first-hand experience with a very large dataset linked to traditional toxicology outcomes. The importance of this is to know that we are going to be developing guidance in terms of the optimal and minimal content and format for the submission of microarray data, and looking at this database has definitely provided us with a very, very good experience as to how they look and the things that we should consider important. So, as I mentioned, Iconix continues to provide training and support in the area of QA/QC, as Kurt mentioned this morning, and analysis of the data across multiple gene microarray product platforms, and the derivation and validation of markers or toxicity and mechanism

from integrated chemogenomic datasets.

[Slide]

Finally, I would like to tell you about a collaboration with Expression Analysis and Schering Plough. This is to develop a mock submission of microarray data, and the data will be provided by Schering Plough.

[Slide]

The objectives are to provide a suitable framework in which to augment, reduce or further define a potential list of recommendations; to contribute to the development of consensus around the specific elements of applicable recommendations within the context of a mock submission; and to contribute to building and refining a process in which microarray data may be submitted to the FDA.

[Slide]

We met with Expression Analysis back in May for concept definition and refinement of scope. We are expecting a pilot submission in July and a completed mock submission by October. This should give us a very good experience as to the details that we need to sort out in order to receive microarray data.

[Slide]

The areas to be addressed during this process of receiving this mock submission of microarray data are laboratory infrastructure, data management, study-specific array performance, experimental design, pre-processing and statistical analysis methods, as well as the interpretation of the results.

[Slide]

For the purpose of this presentation I just want to focus on the data management aspect. It is to attempt to sort out things like data files and file structures, the variables and their definitions, and how to link all this information or microarray data to other databases such as histopathology or clinical chemistry.

[Slide]

I should tell you that the first thing we want to do is just to look at the infrastructure that is currently in place. What we did was we looked at what we have. There is a guidance that was published in January of 1999 providing regulatory submissions in electronic format. Specifically, this guidance says that animal line listings can be submitted as datasets. So, animal line listings that you would provide on paper or in

PDF format may be provided as datasets. So, each domain should be provided as a single dataset.

[Slide]

The guidance goes ahead and gives a list of recommendations. I won't go into a lot of detail, but just to mention some of the salient points, such as each dataset should be provided as a SAS transport file. The size should be less than 25 MB per file, not compressed. There are some specifications about the data variable names and the description of these data variables and the labels. Data elements should be defined in definition tables. Each animal should be identified standard a single, unique number for all the datasets in the entire application. The variable names and codes should be consistent across the studies, and the duration of treatment should be provided based on the start of the study treatment.

[Slide]

This is an example of a dataset and data elements as stated in the guidance. What I would just like to point out is some of this--variable name and it is stated that it should be eight characters. The label should be very descriptive

of the variable. For example, here, lab test is the name of the variable and it would include any other variable, such as clinical, chemical or hematology or clinical science.

[Slide]

This is an example that tells you what the histopathology table should look like. For example, the name of the organ and then the different findings, macroscopic findings and microscopic findings, should be defined after that.

[Slide]

So, we have something in place in order to submit datasets electronically. However, so far this does not include anything on how to submit microarray data. However, back in January there was a notice in the Federal Register on a pilot project for nonclinical datasets. Dr. Randy Levin actually told us a little bit about the CDISC project. This pilot project is part of an effort to improve the process for submitting nonclinical data. Eventually, FDA expects to recommend detailed data standards for the submission of nonclinical data.

The FDA received recommendations for a standard presentation of certain clinical data from

the CDISC and CDISC is currently facilitating the work on similar standards for nonclinical datasets. So, now what we have is some infrastructure and we have an initiative going on, which just points out that this is a very opportune time to try to get these issues resolved.

[Slide]

So, what we did, we went ahead and compared our current infrastructure to some of the mechanisms being proposed outside. So, we compared the CDER guidance to the MIAME-Tox proposal. I should mention that this is by no means an exhaustive comparison but it is just to point out and highlight some of the similarities and disparities that we currently have, again emphasizing that this just points out that it is an opportune time to try to get these issues resolved and addressed.

For example, the CDER guidance paradigm appears more comprehensive with less restrictive vocabulary. For example, the CDER proposal treats LABTEST as a variable, while the MIAME-Tox proposes a field for each possible clinical chemistry test.

Again, what this really tells us is that the CDER guidance is actually more malleable and at

this point will be able to accept MIAME-Tox formatted data. So, if there was consensus that this would be the best way to get the data formatted, then the agency will be able to accept such data.

The MIAME-Tox collects information on in vitro experiments, whereas the agency generally does not receive line listing for pharmacology data. This goes back to what Dr. Leighton was telling us about a little bit earlier, that the requirements for the submission of data that is pharmacology and toxicology are different. For example, line listings are required for toxicology data and are not for pharmacology. Thus, the CDER guidance currently doesn't have a mechanism to accept pharmacology data because it is typically not submitted as line listings.

On the other hand, in a typical toxicology study you generally have pharmacokinetic assessments and MIAME-Tox at this point does not collect information on drug plasma levels. So, these are just some of the differences, very overall differences and similarities but mainly what it points out, again, is that now that we have initiatives going to standardize the nonclinical

terminology, as well as initiatives to figure out the best way to collect a standardized database--that this will be the best time to try to get those two things together and make them compatible.

[Slide]

I am just going to mention some considerations for the submission of microarray data. Based on what I just told you, it seems that it would be useful to have sponsors provide annotations to nonclinical data containing array information by following a guidance-compliant format. That would be with the disclaimer that the guidance may have to be extended to include how the array data may be submitted.

This is, again, something to consider, that is, to include the following files. So, the raw data files post image analysis, and in the case of the Affymetrix array data that would be the CEL and the CHP files, linked by animal identifier; and to include a summary report to describe any normalizations, data processing, and/or statistical analysis, basically how conclusions were derived.

[Slide]

Let me tell you a little bit about the

thinking behind perhaps having sponsors submit these raw data files post image analysis. Here is a table that presents what these files mean, particularly for the Affymetrix data. For example, in this case we would perhaps be asking the sponsor to submit the CEL file, which basically can be used to reanalyze data with different expression algorithms but it basically gives it to you readable in any type of text editor. So, you would have to be able to generate data tables that would be suitable for review purposes. The CHP file in this case would quantify and qualify the transcript and its relative expression level.

So, the question is how about this DAT file? It is 40 MB. It is raw data. At this point we are leaning not towards the submission of this specific file. Some people argue that one of the reasons why you might want to have the DAT file is because you would be able to address issues such as this.

[Slide]

As you can see here, this just shows a defect in this chip, and by looking at this image you would be able to assess that. However, I think we can probably come up with some other ways in

which you can get this information without having a 40 MB file submitted to the agency, perhaps a picture in a PDA format or just the information from the CEL file, or come up with some QA/QC matrix that would allow us to determine the appropriateness of the experimental setup, in this case the chip integrity.

[Slide]

This is just to give you an example of what a probe detection report would look like coming from a CHP file. Again, since this will be able to be modified in any text editor, the tables might look different depending on how the sponsor would like them to look.

[Slide]

So, these are suggestions for submission of array data. By evaluating several submissions we can gain understanding of the fields and issues that need to be reconciled for database purposes. This proposal works with the current guidance. It does not create any additional burden for the sponsor and leaves the possibility of an in-house database creation.

[Slide]

With this mock submission data, what we

are trying to do is sort out the details as to how the data should be submitted, what it should look like, and it also would give us an idea of the things that we need to consider in order to have the best infrastructure to receive this data.

I hope that with this presentation I have given you a flavor as to the main initiatives that are currently going on here, in CDER, in order to prepare ourselves to really understand the field of pharmacogenomics and the regulatory considerations stemming from the development of toxicologies. Thank you.

DR. KAROL: Thank you very much. What we will do now is have questions for any of the presenters, then at 2:30 I am going to turn the session over to Dr. Sistare for him to ask questions of the panel. So, now any of the papers are open for questions. Yes?

DR. SISTARE: A question for Bill Mattes. Bill, one of the fields that didn't come across on one of the visuals that you had was histopathology. What is the current thinking? What is the current status really of the MIAME-Tox menu and choices with respect to being able to pick and choose the descriptors you need for the histopathology? Is it

felt it is robust enough, it is adequate? Do you feel that you have got the consensus of the pathology community and professional societies? Is there some work that needs to be done there to sort of get a better feel that we have the consensus; we have what we need at this point in time?

DR. MATTES: No and yes. No, you didn't see the histopathology. I was trying to keep slides to a minimum and it is always a question what you put in and what you leave out. In the case of histopathology, that was an interesting dynamic we went through. We had considerable debate on what to do. Histopath was obviously collected at numerous sites originally, yet, when we sort of met as a group to discuss how to handle this--we had Roger Brown from GlaxoSmithKline sort of enlighten us, those of us who had not been so up close and personal with pathologists. He enlightened us that, you know, if you have two pathologists you will have three different opinions so he encouraged us to take the approach of having all of the data reread by one pathologist.

So, what we did, we were having Peter Mann at EPO read it and capture it in an EXCEL spreadsheet. It has drop-down menus and controlled

vocabulary. He kind of agreed to it and the nomenclature was basically ripped off from NPT. So, we are in the latter stages of capturing that data. There is good and bad to this approach. The good is that for this particular dataset we will at least have consistent histopath. We haven't entertained the thought of trying to see how that correlates with the previous histopath that was done, obviously not collected electronically, but that is the status.

Now, in terms of how does this jive with the rest of the histopath community, you know, I certainly don't want to die on that hill. I know that is a tall order, to harmonize that nomenclature. I am hoping that in this exercise we might be catalyzing some movement along those lines. As I say, the other thing would be to capture all the separate histopath readings that were done in the individual companies and sort of run an "ooh, what did you think" comparison. But for the purposes of this dataset we had one pathologist read it, or we are having one pathologist read it and that nomenclature is pretty similar to the NTP.

DR. BROOKS: I have a question for Kurt

Jarnigan. A number of the speakers spoke to the importance of experimental design and I think for this technology or most genomics-based technology that is critical. However, you were the only person that provided a number as far as replicates in experimental design goes, and I was wondering if you could go into more detail with respect to your biological replicates of three and whether or not that is something that should be limited to in vitro studies or can be expanded to in vivo studies, and I guess speak to how you arrived at that number and expand on that a little, please.

DR. JARNIGAN: Those were designed to be minimum study sizes. Those are the minimums that we find useful, mostly because that is the minimum you can do any useful statistics on.

DR. BROOKS: But let's say you are looking at human tissue, still a minimum of three irrespective of the control for genetic diversity and some of the other factors in your models?

DR. JARNIGAN: Well, a minimum of three but, yes, probably in those settings--I can only speculate as I have no personal experience with human tissues derived from patient samples, but I would speculate that you would need more than three

to derive any statistical power of any kind in that setting. But for the case of animal studies, which we have done a lot of, I can say that three is very, very good and in a good lab with careful quality control it would be adequate to cover most major toxicological and pharmacological findings. Clearly, for some of the more idiosyncratic findings, yes, you will need more than three to cover those and in some specific experimental case you probably would need more. But for your average run-of-the-mill toxicological findings or the average run-of-the-mill pharmacological findings three will do if the experiment is done carefully.

DR. BROOKS: Do you find that increasing your number of replicates will increase your sensitivity depending on what you are looking at? Or, does it not make a difference at this point?

DR. JARNIGAN: We have only examined between three and six, to answer that question. I haven't gone beyond six but it looks like we are approaching an asymtote pretty quickly and beyond six you don't really get much additional sensitivity. In theory, it is a square root kind of function so you quickly get to a point of diminishing returns in that kind of a situation.

DR. QUACKENBUSH: If I could actually add to that, I think part of the answer to your question depends on what the goal of the experiment is and how you want to do it. There are actually two places in the literature where you can find discussions of this to some extent. One is a paper published by Gary Churchill in CHPing Forecast Supplement to Nature Genetics where he talks about the value of biological replication. Probably a better reference is a paper by Rich Simon. I don't have the journal citation at my fingertips right now. [Simon et al., Genetic Epidemiology, 23:23-36, 2002] I can pull it up on a laptop if you like, but he actually introduces a power calculation for microarray experiments where he goes through and looks at the level of sensitivity you want to approach and the degree of biological replication that you need as a function of the variability in your assay.

So, while I think three is a good starting point, you really have to be much more careful and much more proactive about doing the up front work to estimate what the inherent variability is before you decide on a certain level of replication to reach a certain goal in sensitivity.

DR. BROOKS: So, one could establish a guideline based on the question or the model as to how many replicates would be acceptable for a study so you could properly evaluate the data.

DR. QUACKENBUSH: Exactly. I think what you need to do is look at these power calculations and sort of validate them, and then use that as a standard.

DR. BUSH: I guess what I was getting at is there need to be multiple different things; there can't just be one design.

DR. KAROL: John, is that reference on your slide? This might be a very good time to announce that all of the slides will be posted to the web site so that it will be on the web site, John. There is no need to get it now.

DR. QUACKENBUSH: It wasn't actually there.

DR. ZACHAREWSKI: While we are waiting for that, I was wondering if I could ask Dr. Rosario to talk more about the Schering Plough collaboration. Is the source of the data part of the ILSI-HESI effort or is this a separate effort altogether?

DR. ROSARIO: No, it is a separate effort. The data provided by Schering Plough is not from

the ILSI effort. It is an independent dataset from a compound and they have some microarray data linked to toxicology parameters but it is just an independent dataset.

DR. ZACHAREWSKI: So, it is not just the microarray data, it would be microarray data and all the other supporting IND data that is typically submitted?

DR. ROSARIO: No, no, no. I think not in the context of an IND; it is independent of that. It is microarray array linked to some toxicology parameters, but not within the context of a pooled IND. Basically, the point of that is to sort out exactly how the data should look, what components should be submitted and, you know, sort out variable names and the details of are we able to actually receive the data with our infrastructure, and things like that.

DR. ZACHAREWSKI: So, there will be, like, clin chem and histopathology and all the other nasties and goodies?

DR. ROSARIO: Yes.

DR. ZACHAREWSKI: So, will there be a report about that?

DR. ROSARIO: Sorry, will there be a what?

DR. ZACHAREWSKI: A report.

DR. ROSARIO: Yes. I didn't go through all the different statements in terms of the deliverables. We have a report that should be submitted, yes.

DR. LEIGHTON: With regard to the question of variability, I think it is interesting or instructive to point out that about three years ago there was a very important paper, I believe, in Cell by Yu, et al. from Rosetta Informatics where they were looking at microarray data from a particular strain of yeast that they were experimenting on. In order to make sense of their experiments and get a handle on variability--this is in one laboratory with one sub-strain of yeast--they did something like 50 or 52 controlled cultures to get a handle on variability. Then, once they were able to identify about 80 or 90 genes that varied tremendously in their controls and tuned these out, they then were then able to make sense of their experiments. So, I have become a little concerned actually when people talk about maybe three as the number for mammalian studies.

DR. JACOBSON-KRAM: One of the issues that appears to be quite controversial is the issue of

whether or not studies need to be conducted under good laboratory practices. So, I would like to perhaps discuss this topic and say that any data that is conducted as part of an initial safety assessment, if it is pivotal data, then that should be conducted under GLPs and all other data do not need to be so conducted. We heard a lot about data integrity, data quality going on. It seems to me that good laboratory practices could help this process. I would like to perhaps throw this out for a question for discussion.

DR. KAROL: Any response to that?

DR. JACOBSON-KRAM: Has any vendor tried to validate their system for GLP? I would be pretty surprised. Kurt, do you know anything?

DR. ZACHAREWSKI: Kurt, were your studies run under GLP?

DR. JARNIGAN: No.

DR. SISTARE: I would just mention that the Expression Analysis does perform this function as a service for sponsors, and they are striving toward that end. We are actually trying to hold them back a little bit, saying we don't have to achieve GLP status at this point in time. But they are striving to get there. So, I am seeing efforts

in that direction to do that, but for our purposes, we indicated we don't have to achieve GLP status here. You can specify however you want to the first part, the laboratory parameters that they are following, but they are doing things GLP-like.

DR. KAROL: Are there any other questions? If not, I would like to turn it over to Frank.

Questions to the Subcommittee

DR. SISTARE: We have had a pretty full day. Our attempt, our goal here today was to bring all the committee members up to speed, up to the same level playing field and, at the same time, speak to our outside constituency as well. What we have here is an opportunity to get open public discussion, open public transparency with respect to where the agency is at this point in time in our thinking and in our goal setting.

I think as you can see from what we have done today, we have brought everybody up to speed with respect to where the experts out there in the real-world are in terms of the technology providers, in terms of trying to develop standards, in terms of sponsors, how they are using the data. We have heard excellent discussions from within the agency on what we are trying to do to adhere to

existing standards with respect to electronic data submissions, the kind of playing field boundaries we have to stay within so we don't have to start all over from scratch and create something that creates a lot of havoc in the field. And, we have brought you up to speed with respect to everything we are doing internally as well.

We don't want to be perceived as being way out there and trying to force a future. What we want to be perceived as is as enabling and allowing whatever the best future is for all of us to evolve and to do things a better way. So, that is really what we are trying to do here. FDA's goal is to work as compatibly as we can with our constituency out there. Our constituency is both the American public in terms of assuring the best drugs get to the marketplace, as well as the sponsors who we are highly dependent on to develop these drugs and to bring these drugs to market. So, they are as much our constituency as the American public. We want to work as closely as the regs allow us to, to enable some preferred future and we have to define what that preferred future is.

With that in context, I want to pose these questions. I am just going to go through all of

them, all three of them. We have an hour for discussion and I think the rules are that only the people at the table can comment on these questions. I apologize to those in the audience but these are the playing rules. So, I will invite a lively discussion from all the participants on the committee here. I will go through the questions and I will just invite all of the participants on the committee to dive in on any particular question that excites them the most but let's try to cover them all if we can.

While most data from genome-scale gene expression experiments are incompletely understood, at the same time much of these data are considered valuable. I think each and every day, as we have heard, there is exponential growth in the realization of the value of the measurements of these transcripts. So, it is a rapidly growing curve that we are on. Reluctance, however, has been expressed in incorporating these endpoints into routine pharmacological and toxicological investigations.

The questions are, should the FDA, Center for Drug Evaluation and Research in particular, be proactive at this time in enabling the

incorporation of such study data into nonclinical phases of drug development and in clarifying how the results should be submitted to the agency? What should present and future goals be for use of the data by CDER? What major obstacles are expected from incorporating these data into nonclinical regulatory studies?

Second question, concerns have been raised about gene expression data reproducibility across laboratories, across platforms and technologies and over the volume of data generated from each experiment. First of all, is it feasible, secondly, reasonable and, third, necessary for CDER to set a goal of developing an internal database to capture gene expression and associated phenotypic outcome data from nonclinical studies in order to enhance institutional knowledge and realize the data's full value?

We have had a few submissions of microarray data. They have come to us in paper format. I think we have heard a number of speakers today indicate that that is a pretty difficult way to get any really useful information out of the full dataset. So, the question is should the data come to us electronically in a format that we can

archive and use and learn from?

The third question is concerns have been expressed over reanalysis and re-interpretation of large gene expression datasets. You heard Lilliam say that the CEL file would be a nice file to be submitted. The CEL file does allow reanalysis of the data. Affymetrix data analysis has gone through an evolution from a number of different ways of doing that and we see publications coming out at least once or twice a year on another way of analyzing data. So, if the CEL files are submitted, that would allow that kind of a process.

Is it advisable for CDER to recommend that sponsors follow one common and transparent data processing protocol and statistical analysis method for each platform of gene expression data that would be submitted but, at the same time, not preclude sponsors from applying and sharing results from additional individually favored methods? This would at least allow one beginning, starting level playing field.

What specific advice do you have to us for clarifying recommendations on data processing and analysis, as well as data submission content and format? Our goal over the next six, seven, eight

months is to take your advice and to work from this as well as our experience from the mock submission data and from our own experience from working with gene expression data to come up with a draft guidance that will be used as a template, if you will, for sponsors who choose to--we are not in any way specifying that sponsors have to generate microarray data, but if they choose to generate data and as upper management works out the details of whether data need to be submitted or not; if the data need to be submitted, whether it goes into--I will use the words safe harbor, I am not supposed to use that word--safe harbor or non-safe harbor. The question is how should the data be submitted to us.

So, we are not going to focus on those bigger issues that will be worked out in dialogue with PhRMA and will be handled at a much higher level, but the technical issues of how the data could and should be submitted to us is really what we hope to clarify for those sponsors who choose to and wish to submit their data to us.

So, I leave those questions out there for people to dialogue on. I guess I should just step back and just let you dialogue.

DR. GOODMAN: Well, I would first like to say, Frank, I congratulate you and your colleagues here in terms of wanting to be proactive. It is very, very important. But I think that I would like to make just four points.

I think that toxicogenomics has a bright future, but I think that there is a possibility to short-circuit this by being too prescriptive at an early time and we are, indeed, at an early time.

My suggestions would be to permit sponsors to supply their data as they would write a paper for a high quality journal and allow each to do it, and do it in a scientifically solid, comprehensive and defensible fashion. I would not move to set standards at this time. I would try to shy away from fixing in stone a database now because I am concerned that fixing the database now could then limit the ability to be expansive in terms of the experiments because the experiments may then be done to fit the database rather than following the science.

The other thing that I frankly find a little bit disturbing from the speakers and from my general reading is that in the majority there seems to be a tendency, although no one explicitly said

this, that the larger the number of genes on the array the better and if someone has 15,000 someone should try for 20,000 or 25,000 or 30,000. With all of the difficulties we see in terms of analysis and reproducibility etc., maybe there should be some encouragement to focus on smaller subsets of genes and, in a sense, to start walking before we start running. Thank you.

DR. KAROL: Tim?

DR. ZACHAREWSKI: I would like to disagree with my esteemed colleague. I think it is important to provide guidance and that those guidelines can change as we become more knowledgeable in terms of the structure and the format of the data. I think that if it is 15,000 genes or 30,000 genes it doesn't make that much difference in terms of the analysis.

Interpretation is a different story and what I would really encourage is that with these mock submissions it comes as close to the other required information as possible being provided as well because I think it is going to be that other supportive toxicological data that is going to put that gene expression data into perspective, into biological context. That is key. It will not only

help in terms of making sure you are not chasing insignificant changes in gene experiments, but it will also have significance in terms of providing some kind of direction of what are the significant changes in gene expression and, as NIH likes to call it, phenotypically anchor those changes as well.

I can't remember what other point I wanted to disagree with. Do you want to share that again?

DR. GOODMAN: Just leave it as a general disagreement.

DR. ZACHAREWSKI: Yes, we will continue this on the plane home.

DR. HARDISTY: I feel that the FDA should be proactive in any initiative like this. My concern is that it may be a little bit premature to incorporate these into routine nonclinical studies and make them a requirement. I hear there is a lot of need for standardization in the way the tests are run, the protocols, the nomenclature. So, it seems like it is very early in the process and it may be that on a drug by drug or class of drug basis that data may be very useful in helping in risk assessment, but in most instances it is going to be part of the evidence to support an overall

decision based on more standard toxicity studies.

I think though that this is the time for FDA to get involved in it when it is early in the process so that you can help lead it. Right now I see that there are two or three groups almost progressing in parallel and there is a lot of overlap between those groups in nomenclature, protocols and things like that. It is going to be important to have some coordination between those groups.

I just might mention a little bit about nomenclature as a pathologist. It seems like there is a lot of discussion about pathology nomenclature. I realize that on this first study one pathologist is going to reread all the important target tissues. It may be a little impractical down the road if studies are submitted to the FDA to have one pathologist reread all the important target tissues. Now, if you do have one pathologist and he uses one set of nomenclature such as that Dr. Mann is going to use the TCMS nomenclature, the TCMS nomenclature in Dr. Mann's hands will be fine but it is a list of words; it is not a list of definitions. So, another pathologist can use that same list of words and define them

more in line of his thinking as far as those words go. So, I think that before we decide on which nomenclature is accepted or is used, it may be good to get a group like the Society of Toxicologic Pathology or them in conjunction with maybe the Society of Toxicology to look at this problem of nomenclature and try to tie these changes in gene expression to biologic changes in the tissues. It is something that I know some of those organizations will enjoy working on and will probably do a very good job.

DR. BROOKS: I agree that FDA's involvement in establishing guidelines now is a good thing and that it is not going to hinder or inhibit the development or the use of this data. In fact, it may enhance it. Because of the fact that there are so many different people, using so many different technologies, doing so many different things, without guidelines toward a specific goal it is going to be much harder for people to achieve that goal. I think even independent programs, whether it is academia or industry, are struggling with how they should be doing things. So, some guidance from the right perspective I think will be very helpful and I

think the FDA can be very constructive in that and, as we learn more about the data and its ability to be more informative for these applications, those guidelines can become more rigid but right now they can remain flexible.

With respect to the number of genes and the data overload, there really are, you know, two schools of thought and I think that some people that started working immediately with specific arrays are biological questions and if you make an array where 99 percent of genes on that array change as a function of your model, data analysis becomes an even more difficult task. Biological is broad; the arrays are broad and some of that information that may not be used specifically for biological inquiry is very important for normalization and for understanding the systems that you are interested in. So, I think data analysis and the mathematical problems associated with data analysis will continue to evolve.

But as Dr. Quackenbush stated, the fact of the matter is you really do need to define your question in order to be able to use this technology effectively, and what the FDA has here with respect to what they are interested in, toxicology, can be

a very well-defined question. If they can define their question, they can use this technology probably better in some instances and I think that the question is here; it is just how well we can define it.

With respect to building a database, I think databases are good. We create them; lots of people create them. I think that if the FDA wants to start to look for its own development and for its own information, not necessarily to hold that information against sponsors but to use it to continue to develop their question and their guidelines, having that data at a raw level is going to be important. So, as new mathematical analytical models are established they can use them to their benefit and not necessarily to the detriment of their sponsors. Data analysis is the one thing--you know, the technology has allowed us to accelerate the development or the creation of data tremendously. However, we really do in some respects lag with respect to what we can do with all of this data and being able to look at thousands of genes at a time and how it relates biologically. The guidelines I think should focus on some of the technological variability which

allows us to focus on the biology. But from an analytical standpoint for biology I think the FDA needs to be involved in what analysis it feels necessarily is important or what it will run or expect to see, and that is probably the most difficult question that I think faces some of the guidelines that need to be created.

DR. WATERS: I would like to just pick up a little bit on Dr. Hardisty's comments and try to move them into the realm of toxicology. I think we are really at an early stage in understanding how to interpret molecular expression data in terms of toxicology. I don't think we have put molecular expression on toxicologic pathways yet. I think we are just beginning to do that. I think we need to understand those pathways in a molecular expression context.

As we move towards that kind of an endeavor and as we move towards building databases we very definitely need to develop ontologies in the toxicologic domain as well as the pathologic domain. Those ontologies will be critical in common understanding, common database query capabilities in the future.

So, I do believe there is an important

need for consensus building and for international efforts in doing this sort of thing. The MGED Society has made an important start. There was a contrast between MIAME-Tox and the efforts that are ongoing at CDER. The MIAME-Tox effort is just the beginning of an attempt to put forth a potential guideline in the toxicology domain. I think there needs to be participation and there has not been participation thus far in clarifying that guideline. So, to me, there is a lot of room for us to define the domain of toxicology, to separate that domain to some degree from the domain of pharmacology to really understand what we mean when we talk about toxic effects in a molecular expression context.

In order to do that, we do need a database. The question is do we really know how we want to build that database at the present time? Do we really have enough standards? Do we really have enough ontologies? These are things that I think are important to consider. Thanks.

DR. KAROL: We have remarkable agreement that we really should link molecular expression and toxicology and pathology, and that we shouldn't be too restrictive. But I would like to hear a little

bit more discussion about this database and what you think should be involved in creating an effective database. Frank, do you have comments?

DR. SISTARE: I was just going to say one thing. I don't know if this is one of the things that Tim was forgetting with respect to what Jay had mentioned, but Jay mentioned something along the lines of we ought to model data submissions to the FDA along the lines of the way a paper would be put together and submitted for publication. But I think as John Quackenbush pointed out, those journals are requiring the full gamut of gene expression data derived from those experiments to be submitted into a database. So, that is routine now. Those journals are not publishing data without people having documented that they have submitted the full gamut of gene expression data into a database.

So, it seems like that is becoming the standard, the societal standard, if you will, for supporting the conclusions of a well constructed microarray gene expression experiment, that is, full disclosure of the data that support the conclusions of the paper for the inquisitive scientists who look and evaluate on their own.

So, your question, Meryl, I think is spot on and that was one of the first questions. You know, format defines utility of everything, or the shape of something is defined in utility of something. If we ask for paper submission, it is only going to be useful for that particular context which the paper is being submitted to support. That is all it is going to be useful for. If the data is submitted electronically it now expands the utility of that information.

So, I think that is the first fundamental question we have to establish. FDA is moving toward electronic data submission. It just happens to coincide with the fact that now we are getting 10,000 data points on an experiment and the only way you can really make sense of that is if it is submitted electronically. You know, we are establishing the first, fundamental question, which should FDA ask for the data to be submitted electronically? The first question is, is that a reasonable request? Once we have established the answer to that question, if the answer is no, okay, we can go home but if the answer is yes--maybe we should just ask that question first.

DR. ZACHAREWSKI: Just to follow-up, you

said that you are going towards electronic submission. That means that minus the microarray data you already have a database established to capture all that information. Is that correct?

DR. LEIGHTON: We have to be careful here in distinguishing between electronic submission of paper data versus submission of electronic data. I think the way we would be moving would be submission of electronic data so that it is truly searchable and can be searched across submissions.

DR. ZACHAREWSKI: But would you store that within a database housed within FDA?

DR. LEIGHTON: I think ultimately, because of the proprietary nature of the data, we would have to do that. I doubt that it would be public.

DR. ZACHAREWSKI: So, that is the plan, to develop a database to store that data only for FDA use, period?

DR. SISTARE: Well, I think the initial plan is to enable submission of electronic data in a way that it is very easy for the reviewer to move around that data and to pull things together and pull it into programs to analyze the data electronically. So, that is really the visible rationale for doing it. By the way, once you do

that, now you can create a database and I think it would be unwise not to. I am going to ask Randy to address the question. I think you are asking sort of the status of things right now. There is not a lot of electronic data being submitted to my knowledge.

DR. ZACHAREWSKI: Yes, there are two questions, the status and will the system that you have allow you to query across submissions?

DR. LEVIN: We are working on the tools that will help us analyze that but we have found that we are going to have to put that into a database for those tools to work efficiently. So, we are aiming toward a database that we put the data into. If we develop a common terminology, then we can potentially look across studies.

DR. ZACHAREWSKI: You mean like the MIAME-Tox ones?

DR. LEVIN: Well, for example yes. The thing that we are focusing on first is the structure of the model, so not the terminology. We need both to be able to look across studies.

DR. ZACHAREWSKI: The only other thing I can say is that it sounds great but it won't happen in my lifetime. So, when is this actually going to

be in place? That is the other thing. I think that is going to be another major impediment because these are not small undertakings and I am sure you appreciate that.

DR. LEVIN: Well, we have gone pretty far with the clinical data to define how we can transport the information that we need for making our regulatory decisions. We have a pilot project for both the clinical and nonclinical data so we are hoping that we start to receive some of this data in from our pilot this year and to test the model and see how good it is.

DR. ZACHAREWSKI: So, that means that you could take this model and just add on to it a subsystem for microarrays. Is that the plan?

DR. SISTARE: Yes, and I think what Lilliam described is right now--we have a document out there that says here is a way that you can submit electronic data if you want to, right now. I think the status is that we just haven't received that many electronic data submissions but it has been an option for sponsors to do at this point in time. We are not making them, we are not requiring them to but, again, allowing and enabling. So, now within the context and the boundaries of what we

have established, if a sponsor chose to adhere to the MIAME-Tox guidelines that are out there they would be compatible. There are just a couple of small things where we may have to wrinkle out some things but otherwise they are compatible. MIAME-Tox is more prescriptive, if you will.

DR. LEVIN: Actually, we have had some success with carcinogenicity data and we have been receiving that electronically for a long time. More recently people have been following the standard that was published in the 1999 guidance so that has been pretty successful.

DR. GOODMAN: I think in terms of doing things electronically it really is sort of a no-brainer these days. We should move towards doing more and more, if not everything, electronically. When I said to submit like a manuscript, obviously there would be appendices that would include full data.

My concern, again, is that at the status that I see toxicogenomics today I think to start putting in place a proscribed database would be less productive than over the next few years letting the applicants submit their data in a file form and then take and see what might be the best

of these, rather than start--once you start putting something into guidelines--I hear you in terms of that it can be flexible and it can be changed, but it gets much more difficult. It gets difficult to start changing once you have guidelines.

I just wonder out loud whether the notion of comparing and sifting and sorting of these database publicly is really something that is realistic. It is my impression that you would be dealing basically with proprietary data and that this would not be that readily available. Maybe there is a certain time span when it does become available. But the point is that in order to really move this field forward it is going to take, I think, industry buying into it, and in order to do that it has to be where you see that it is going to be productive in terms of help, not only help make better decisions but help in terms of working with Food and Drug Administration. So, again, I just think early on the less prescriptive and the more working as partners, I think the more productive everything will be.

DR. ZACHAREWSKI: No, for that session. The problem is that if you don't set up some guidelines, when you do finally set up guidelines

you will lose that information because it will be very difficult, if everybody submitted their data in a different format, to then reformat, you know, what you have collected for the last five years and put it into the proper format to put into the database. If you only have two formats being submitted it is not so much of an issue. If you have 15 or 20 or more, whose responsibility is it to reformat that so that it is acceptable into the database?

DR. GOODMAN: Do you have a crystal ball at this time to start setting up these databases? Why not just see how the information flows for a while and then try to revisit this issue?

BROOKS: Maybe the definition of guidelines is where we are getting hung up with respect to the kind of data to be submitted. Maybe if we start with more simple things as formatted data, as someone said CEL files or raw data versus processed data. Raw data gives you the ability, as new analytical tools for what you want to do across databases come out, the flexibility to do that without restricting you to guidelines with respect to other ancillary information that goes along with it so you use maybe MIAME-Tox as a standard and say

we are going to take raw data. After you start taking that data and working with it, then you can refine or establish specific guidelines about information that is more pertinent to what you are trying to do. But I think the form of data is probably the most critical right now.

DR. SISTARE: Yes, I would add one of the things that Randy pointed out to me and I should have mentioned earlier too, and that is what is important here I think is to specify the transport file, as you point out, the format that you want the data to come in. Then, you can modify that and change that any way to fit a database.

The one place where it does get a little dicey is when you start specifying ontology, words and vocabulary and things like that. If you do that up front, that may be difficult and you may lose some aspect of the flexibility of the use of that information if you don't do that up front but I hear what you are saying, if you are a little too prescriptive and the Society of Toxicological Pathology hasn't quite developed a consensus on the best definitions of the terms.

FDA can maybe proceed judiciously and carefully along that line but are we getting the

general gist that this is a wise endeavor for us to go down; this is a path we should be going down in terms of setting up and preparing ourselves in a way to receive the data, that it could be useful and populate a database without being prescriptive?

DR. GOODMAN: I think the answer is yes.

DR. KAROL: Randy, did you want to say something?

DR. LEVIN: Yes, I think Frank was saying that from our experience and with the clinical data--many things that you were just bringing up--we can define the transport, just the information how to communicate with each other. Our database may change over time but we are hoping that the transport information would stay the same so you would have that stable.

Another piece that might be interesting is the annotated ECG waveform data. We were talking about receiving that in an electronic format. At first we might not have the full database but we would have the standard of how to receive the data. Then eventually, once we got everything worked out, we could have it put into a database. We could take that data we received in the past and put it into a database because it is all standardized.

Then, the other thing is that once we have the database there is a possibility to look at some of that data for research issues beyond just a review of that particular application. So, looking at it and saying is there some way we can monitor drugs for cardiac toxicity because we look at this ECG data. So, it does offer something beyond the initial use, and something you would consider for your work here too.

DR. HARDISTY: I agree. I think it is a good time to probably start a database and it should have some minimal standards. I think that is what you have recommended. If someone wants to go beyond that, so be it. So, it is not really restrictive or prescriptive but there is some minimum that you want everybody to conform to.

The other thing about restrictive nomenclature I think is probably a good thing and not a bad thing, particularly in histopathology or any of the toxicology endpoints. We have been doing toxicology studies for years and we are trying to take the information we get from toxicology studies today and correlate it with gene expression. So, the things that we are seeing in the tissues aren't going to change. We are trying

to correlate those changes with gene expression. So, we should be able to go ahead and restrict the terminology based on what we already know. What we are trying to do is eliminate synonyms in our database so that you can search it without having to worry whether the study was done in England or whether it was done here, in the United States. So, I think that we already have the information there. It is just a matter of setting it down and deciding what you want in your database and how you want to handle it.

DR. BROOKS: One thing that was mentioned in the first talk with respect to the goals--one is to, obviously, find more sensitive or different ways of assessing toxicological assessment. The other is being able to make predictions based on the efficacy of drugs and their toxic events on specific individuals. So, I just wanted to note that without collecting data from individuals or studies that are specific in having that full dataset it is going to be virtually impossible to achieve that second goal. So, having a database is going to help you make greater strides with individual sponsors or academic labs that are trying to achieve that information. It is a much,

much larger endeavor that needs to be at the level of the federal government I think.

DR. WATERS: I would just like to comment that I think the FDA can play a very important role in consensus building with regard to some of the data standards. I am not sure that you have been involved extensively up to this point. I think it would be very good if you were engaged in that activity. The international standard setting effort for databases is very important and, as well, the ontology building efforts that a number of the societies are becoming engaged in. So, I think to become engaged actively in those processes and work towards the evolution of also publicly available data so that there could be a consensus in understanding the way in which one would interpret those datasets would be to your advantage because everybody really needs to get on the same page. Everybody really needs to have a common understanding of molecular expression datasets, not only the regulated community and the regulators but also the other academic members of the scientific community, as well as other governmental agencies.

So, I think as well inter-agency efforts would be laudable at this point and there should be

an effort to extend to other parts of the federal government. So, for example, the National Cancer Institute is also developing large databases and it is also interested in the clinical domain. I think there would be natural synergy to work with them in their database efforts. Similarly, NIEHS is very interested in animal toxicology and is engaged directly in developing a public database in that domain.

The other aspect that I think is important is an international one. I think we don't live in isolation anymore in the U.S. We are definitely a part of the international community and we also have to engage in the international sector with regard to development of standards.

DR. HARDISTY: One of the questions was what major obstacles would you expect down the road. Most of the work that has been done with gene expression and genomics has been done in universities or non-GLP type settings. Not that they are not good studies, but it is a different type of environment than in the regulatory GLP laboratory and validation of the systems that you are using and all those types of things are something that the manufacturers and some of the

people who are doing this work need to start thinking about now, rather than later. If these do become regulatory requirements, then they are going to have to work in the GLP environment. Right now, toxicology may be outpacing the science in that area so it is hard to keep--you don't want to not continue the technologic development but imposing GLP requirements on those people at this point. But if these are going to be used in a regulatory setting, then you are going to have to try and limit those areas.

DR. BROOKS: I think one of the other hurdles you might need to be prepared to overcome, with respect to any time you put guidelines in place, is that you are going to get questions about those guidelines and ask for recommendations with respect to how people are going to do things. So, there was a lot of talk about biological replicates, and experimental design and study design. Everybody does things a little bit differently. I think it has gotten a whole lot better over the years with this, but I think that you need to be able to be prepared, given the model and once your question is defined, to be able to answer questions with respect to suggestions. If

we want to generate this data or we want to submit it, you know, what is going to be better, more replicates, less replicates, with respect to our design as these experiments and studies are being built. If you have the guidelines and can't provide some suggestions or information I think that people will be less reluctant to provide that kind of data, fearing that they might miss the mark.

DR. JACOBSON-KRAM: I think it is kind of interesting that the dichotomy that is developing here is the way that we are going to deal with this kind of data versus traditional. For example, somebody submits the results of a carcinogenicity study; you don't ask for the slides. You pretty much believe what the report says and if you are very unhappy with it you can go back and audit it. Here what you are asking for is essentially the equivalent of the slides so that you can reexamine it and perhaps re-interpret it. That is really a change in paradigm for how we have done toxicology in the past.

I think that could also be part of some of the needs in the pharmaceutical industry because basically you say here is carte blanche; go ahead.

Here is how we interpret it; what do you think?

DR. SISTARE: I think part of what appears to be a dichotomy there--I think Kurt Jarnigan expressed it well when he talked about the youthfulness of the technology, the youthfulness of using RNA transcript measures as endpoints to link definitively to outcome, as opposed to the maturity of the two-year bioassay and not asking for slides.

We are striking a compromise and what William proposed is we want a suggestion, a consideration for discussion and for some input in terms of what our thinking here is, not to actually ask for the 40 MB TIF image files. That would, I think, be asking for the histopath slides. So, we are asking for something in between, not just the process report but, you know, the data--the data. I think, again, we are asking for the raw output data. Even that is not completely raw because some algorithm has to be applied to get a signal out of background and, you know, you are allowing the experimenter to do that and not questioning that in a sense when you go to the CEL file, intermediate file. So, you are actually asking for number output.

It is a fair question and it is something

that we have wrestled with and had dialogue on, that is, how far back do you go, and I would like to get some feedback and some dialogue here from the experts who have wrestled with these datasets and know the state of the technology. Should we ask for a polished, final expression ratio report, or should you ask for something like a CEL file?

DR. HARDISTY: I don't see it as a whole lot different than what you get on a carcinogenicity study. You don't get the glass slides but you get the individual data and every data point in that dataset. If you get it in a CEL file and you evaluate and your interpretation is different than the sponsor's, they are going to get a letter from you--

[Laughter]

--so I would see it the same way. You are not asking for the microarray, it is the data that they are submitting so you are not going to repeat the generation of the data, which is what you would do if you had the glass slides. You are repeating the analysis of the data, or could repeat the analysis of the data, which you can do with routine toxicology data today.

DR. BROOKS: I think a lot of it stems

from the interpretation of these datasets and I don't think that the problem is going to be with any given sponsor, that you are going to necessarily disagree with their interpretation but when you look at compounds or things within the same class across sponsors how do you interpret each of their individual interpretations if they are all using different platforms, or even if they are using the same platform, even though they have given you their MIAME-Tox standards tell you that their labeling samples quite differently?

So, I think by having intermediate with the absolute raw data to some unprocessed data allows you then the flexibility to potentially compare across platforms and, more importantly, compare applications as to whether or not there is a consistency for those compounds or those submissions. I think in the case of Affymetrix, the CEL file is a good compromise because it leaves you open for different kind analyses you can do to explore the interpretation, I mean within the context of what they are trying to say. If you had some kind of a measure, as William said, that would tell you if there was a defect with respect to image file, and the same can be true for

slide-based arrays where there is a standard background subtraction, and I think most people won't necessarily argue with respect to array performance and then, instead of getting ratios, getting the signal data along with those would be equivalent to a CEL file.

DR. LEIGHTON: I had a question that goes to the question that is on the board here. For the FDA to specify a transparent data processing protocol and the single statistical analysis method, would this be viewed as moving the field forward or being too prescriptive? Or, should this really be deferred until the issues of standard development are more evolved?

DR. GOODMAN: I think it is too prescriptive. Frankly, I think we have problems in terms of making sometimes too many mistakes in toxicology and we don't want to bring on a new technology and make more mistakes quicker. It is not ready to jump in now in terms of prescribing an approach.

DR. ZACHAREWSKI: I would like to agree with my esteemed colleague--

[Laughter]

--if that is worth anything. But I think

this is one of the issues in terms of what data do you get. So, I would say that if you were to try and prescribe a specific data analysis, which one are you going to choose? And, if you asked everybody in this room, they would probably give you at least two opinions. So, there is no prescribed method at this point in time. However, let's say five years from now when there is, you are going to have to go back to each one of those pharmaceutical companies on bent knee potentially and ask them for their raw data files to be able to reanalyze all that information and repopulate your database using a standard normalization or quantitation type protocol.

DR. BROOKS: That is if you don't collect the raw data now. That is what you are saying.

DR. ZACHAREWSKI: Right. But if you do that now you could go back and do that yourselves with respect to the interpretation, not to go back and, like I said before, penalize what has happened in the past but move in a better direction for the future. So, I agree that right now is absolutely not the right choice. Actually, if you guys have a transparent statistical analysis method, I would like actually to take that back with me on the

plane but I don't think that exists at this point in time.

DR. SISTARE: We could name one but you might not like it. I mean, the rationale behind this question is this whole concern about FDA taking a dataset, analogous to what Jerry brought up--and say here is how we are going to analyze the data when we get it; this is what we are going to do with it; these are the rules we follow. I think there is a lot of anxiety if data is submitted to us by sponsors. They may feel that this is the best way to analyze the data. If we don't agree with their approach and we analyze it another way, you know, will the conclusions be markedly different? Probably not, but it is an attempt for FDA to try to be somewhat transparent and to say, you know, at this point in time this is how we are going to look at the data when you give it to us so you might want to look at it that way first too. You can use whatever other way you want, what you think is best, but you might want to do this because this is what we might do. But if you are what you are suggesting is there is just no way we could do that--with Affymetrix we could say, you know, use 5.0 and we are going to use this

approach.

DR. ZACHAREWSKI: What I would do then is I would encourage Dr. Rosario, when she is working with Schering Plough, for them to analyze their data two different ways at least.

The other thing that I would really do is I would encourage for you to approach other pharmaceutical companies and see whether they would do it, and see how they would do it differently. I don't know whether they would go and talk to Schering Plough or not and just copy what they are doing, but I would think that the idea of getting different perspectives from different pharmaceutical companies--you know, you could then merge and pick what you like and ask them to resubmit what you didn't like.

DR. WATERS: I think actually Tim brought out a major point, and if you look at the LCF I think it bears is, that is that in the effort that was undertaken involving 30 different pharmaceutical companies so much was learned by looking at divergent opinions. I think at this point in time we would all be well advised to look at divergent opinions. We just don't know enough and I think that we have an opportunity here to do

it right and, if we do it right then this technology will become established and we will be able to use it and we will have all we want out of the effort. But I think if we push it too far too fast, then it really may backfire on us.

DR. BROOKS: I think that your sponsors now that would risk--risk is a bad word but that would go ahead and submit data of this nature are sort of at an advantage because I think that you are going to gauge some of your interpretation in the analysis based on these submissions and how effective they are and how well they work, whereas if they wait until guidelines are established they might be changing things in a big way. So, I think that by submitting data it has to be clear that you are not going to necessarily change now the interpretation of the data based on your learning curve or based on how it might be used to establish other kinds of tools. You know, the earlier you get in and can justify your interpretation and your model with your data, it might actually become a better established guideline.

DR. ZACHAREWSKI: Actually, I have another suggestion. Why don't you ask the PhRMA companies how they want to submit the data?

DR. SISTARE: We actually have. We have had at least one sponsor come to us and say we have some data we want to submit; how do you want it? I put the mirror up and I said challenge us. You submit the data to us in a format that you think is the best, the most advisable, productive format, but I did share one word with them, an adverb actually. I said electronically. I did say that but I said in whatever format you choose and, you know, tell us how you would like to submit the data and maybe we can get some dialogue on that and give you some feedback. But we haven't seen it yet.

DR. ZACHAREWSKI: But this might be something that ILSI-HESI might want to pick up. I mean, the organization and the structure is there for them to do that since they meet regularly anyway.

DR. KAROL: Frank, I think we have addressed all of the questions.

DR. SISTARE: I think the feedback we have gotten has been really excellent. I really want to thank all of the speakers and all of the committee participants today. This has really helped us and this is a landmark meeting for all of us. As Helen pointed out, this is the first time we have

assembled this subcommittee. I want to thank Meryl for chairing this beautifully, for getting us back on time and for allowing for full discussion of the issues. Again, I think we got all the issues out there that needed to be. We missed Roger; there was a void there. There was one gap there in some of the practical applications of some real live scenarios that we were hoping to get. But, otherwise, I think we got everything on the table. We have achieved our goal of being as transparent as we can. Now the ball is in our court, and we will try to get back to the committee members something in writing within the next six to eight months that captures some of the feedback we have gotten today and allows FDA to move forward.

DR. KAROL: I also want to thank the committee for a very wonderful discussion and just a very exciting topic. I am really looking forward to seeing just how this new technology can be used in an effective regulatory role. So, I thank everybody for their participation, the agency and Kimberly as well. The meeting is officially adjourned.

[Whereupon, at 3:25 p.m., the proceedings were adjourned.]