

Genetics Perspectives on Policy Seminar

“Genetic Biobanks: Deposits, Withdrawals, and Consumer Protection”

December 9, 2008

**National Press Club
Washington, DC**

(transcript has been edited for clarity)

Kathy Hudson: Good afternoon. My name is Kathy Hudson. I'm the Director of the Genetics and Public Policy Center. And I'd like to welcome you to this seminar, GenePOPS, Genetics Perspectives on Policy.

Today's seminar is on “Genetic Biobanks: Deposits, Withdrawals, and Consumer Protection.” I'm going to give a little bit of an introduction and then introduce all of our speakers. They will each speak for about 10 to 12 minutes, and then we'll open it up for your questions and comments.

I'd like to ask if everybody would turn their phones and Blackberries to silent, so that we don't interrupt during the course of the discussion today.

The seminar today is about biobanks. And really what we started to see over the last several years is an increasing number of large collections of biological specimens and health-related information.

Part of this has been driven really by new technologies in genetics that allow scientists to detect smaller and smaller risk factors contributing to health and disease. I'm sure that Teri will talk some about this today. Reflecting the popularity of biobanks and genome-wide association studies, these new technologies to detect weak contributors to disease. You can see that in the scientific literature the number of references to biobank and genome-wide association studies has accelerated over the course of the last three or four years.

Biobanks are popping up around the globe in many countries. In some cases these are national efforts set up by the federal government, and in other cases they are smaller efforts, although still large collections of biological samples.

In the United States, there are a large number of biobanks that are being created. This is a list of just a few of those fairly substantial collections of biological samples and health-related information.

So there's been quite a bit of discussion about the issues raised by biobanking. And I think we'll talk some about those issues and whether or not those issues are new issues or whether or not they're old issues, just magnified because of the scale of some of these collections.

This lists some of the issues that have been talked about in biobanking. And I think that our speakers today will address some of these.

So with that, let me introduce our three speakers. Teri Manolio will be our first speaker today. And she'll provide the context. She's the director of the Office of Population and Genomics at the National Human Genome Research Institute and also serves as senior advisor to the NHGRI director for population genomics activities.

Before coming to NHGRI she was at the National Heart, Lung, and Blood Institute at NIH where she worked on large-scale cohort studies, including the Cardiovascular Health Study and the Framingham study, which I'm sure many of you are familiar with.

At NHGRI she leads efforts to apply genomic technologies to population research. And we'll hear a little bit more about that. She received her M.D. from the University of Maryland and her Ph.D. in human genetics and epidemiology from Johns Hopkins University.

Our second speaker is Joan Scott, who is the deputy director of the Genetics and Public Policy Center. She's a genetic counselor by training, and that raises its head from time to time in the office.

She has many years of experience in clinical genetics and biotech industry and in genetics policy. Prior to coming to the Center, she was the director of a biorepository at Genelogic, which is a biotech company. She has practiced clinically in lots of different practice settings and ran the genetic counseling training program at the University of Colorado Health Sciences Center. She holds a Master's in genetic counseling from the Sarah Lawrence College.

And then our cleanup speaker is Zak Kohane, who is the director of the Children's Hospital Informatics program and is a professor of pediatrics and health sciences and technology at Harvard Medical School. He leads lots of collaborative studies at Harvard that use genomics and computer technologies. And we'll hear about some of those today. He received his Ph.D. from Boston University and his M.D. from Boston University, and did his internships and residency at Children's Hospital in Boston. So with those introductions, I will turn the podium over to Teri. And let me see if I can pull up your talk.

Teri Manolio: Thanks. Good afternoon. Delighted to be here. Joan and Kathy had asked me to give a little bit of an overview of biobanks and cohort studies and sort of the intersection of the two and how they will be used for genomic research, so I probably won't take my full time. But hopefully I will give you some useful information over the sirens -- God knows what's going on out there. You probably know, but we don't.

Opportunities are presented by biorepositories. Biorepositories could be considered to be any fairly large resource that collects specimens appropriate for genome-wide studies.

Typically, they're linked to electronic personal health information or electronic medical records. Very often they're set up in hospitals or clinics where kind of everybody who goes through the clinic is asked to sign a consent form. And if they do, a blood sample is taken and subsequently studied and then associated with medical record information.

There are biobanks or biorepositories in which information is collected for clinical care and then later used in genomic research. The consent provisions of those vary. And in some cases they're very old collections that have had tens of thousands of samples that people didn't want to go to waste, so there's been a lot of debate on the best way to use that information and those samples. That's not what we're talking about here.

As I said, these are increasingly being established by clinics and hospitals around the country and around the world, and they could conceivably really facilitate population-based studies of genes and environment and their influence on common diseases.

"Biobank" in and of itself -- this is actually a Hungarian website that has a fairly nice description of it. The Hungarians wish to establish a biobank as well. So far they have their website, which is cool. It's in English, which is even better for me.

They define it, and so would we, as tissue specimen-extracted DNA, RNA, proteins -- anything that would come from those that could be used for genomic research. They are generally used to study etiology of disease and also drug effects, potentially development of drugs or development of diagnostic procedures or protocols.

Typically they require separate or additional consent. Not always, but usually they do. Tissue samples can be stored in many ways in order to permit multiple analysis in the future. They could be frozen. They could be stored fresh. They won't last very long stored that way, then extracted for various things. They could be paraffin embedded. There are a lot of different options.

There's often a clinical database of detailed health data collected over time. Again, this is usually linked with the patient's or donor's usual source of medical care, so they're as complete as possible. That's an integral component of these kinds of repositories. And specimens and clinical data are registered with a code number rather than a private identifier to maintain to the degree possible patient privacy.

So there's a bit of a continuum between biobanks where you essentially collect the biological samples. You have some baseline data collection. There may be a separate questionnaire. There may not. It may be whatever just happens to be in the medical record at the time.

And then follow-up primarily through medical records but not with contact to the participant, versus cohort studies like the Framingham Study where people come into a study. They're recruited. They're seen often in a typically lengthy examination at entry and then followed up, seen every two years, in Framingham's case, or every three to five years, whatever it might be, but are revisited for medical events and further exposures and things like that.

So you can sort of think of studies that might be sort of biobank-like. On the pink side here, very large, maybe clinic-based, minimal baseline data and minimal follow-up, other than what's in medical record. And those that are more cohort-like are smaller because there are much more intensive efforts.

Population-based, they're representative of the population from which they're drawn. So if you were going to study Washington D.C., I don't know how you would define the population base of that.

But you'd then want to draw people who could be generalized back to that population. Extensive baseline data and in-person follow-up would be characteristic of those. These tend to sort of merge.

I figured you all knew what the U.S. looked like. I superimposed this on a map of U.S. territorial acquisitions. In case you don't remember where the Gadsden Purchase is. Down here in Arizona. I know Jim Gadsden, a relative of the James Gadsden who that was. I like to show it.

There's a repository up in Wisconsin. One of the first, the Marshfield Clinic, develops a personalized medicine program. They have about 20,000 patients so far and are aiming for about 100,000, I believe. Northwestern University has one here in Chicago. Vanderbilt University in Tennessee. Kaiser Permanente. There's several of them among the Kaisers. The largest is the one that's in Oakland. Aiming for three million participants and samples.

Children's Hospital of Philadelphia is the only one I'm aware of that involves children. And they're aiming for 100,000 kids.

Biobanks outside the U.S. now, superimposed on the 2005 world age structure. So if you wanted to know the countries that have 60 percent of their population under age 30, I believe it is, that's Africa. And these have about 30 percent of their populations of young ages. But at any rate, yes, it's quite a difference when you want to understand Africa and the developing world, you need to look at kids.

CARTaGENE in Canada is one fairly large one. deCode in Iceland. Generation Scotland, UK Biobank, 500,000 people there. LifeGene is a Swedish study. There's the Estonian Genome Project, about 20,000. Biobank Japan. The Kadoorie Study in China funded through the UK Biobank and collaborators in Oxford. Taiwan Biobank and the Joondalup Family Health Study as well.

So this is kind of a list of those outside of the U.S. and showing the size, the countries that are involved.

To get access to the data and the samples very commonly is with collaboration, and that was the typical model for these kinds of studies up until about five years ago or so. And over time that has opened up so that people can apply to use data and samples without having necessarily an established collaboration with the investigators. And you can see the ones that currently are showing that they have a controlled access process and a procedure that one would go through in order to obtain samples and data.

And it's expected that many of these others will probably move toward that model, but they haven't yet.

I did just want to mention some collaborative efforts of trying to put these biobanks and repositories together. Probably the biggest one is the Public Population Project in Genomics, the 3PG Program, and that lists a number of them on their website that's shown here. This is just a map of all of their collaborating sites. So there are quite a few. Not all of them are up and running.

And I did want to mention also some of the challenges that might be related to using electronic medical record data, and I think Zak is going to touch on this as well, for this kind of research. Issues related to validity, biases, limitations in the data, completeness of the data, research use by investigators outside a given EMR system.

So how can people outside of Mayo get access to the Mayo data, particularly if Mayo has told Mayo patients that only Mayo investigators will have access, which sometimes happens. How to extract information, how compatible are existing formats. How does one deposit the data. How good are the documentation, the deidentification. And how user friendly is it? Do you have to spend a year sitting at the elbow of the people who put this together in order to understand how to use it, or is it something you can learn from the web or from an online course. There are also issues related to consent, community consultation, that Joan will be talking about.

Recognizing that as these technologies evolve, they do present really special challenges in not only maintenance of privacy, but also what does this information mean, what impact does it have on an individual -- on their reproductive choices, on a variety of things.

And recognizing that consent for research use may not be adequate for clinical specimens, and how does one deal with that? And also issues related to widespread data sharing often have not been assessed very well, particularly in communities, where very often findings are sort of generalized back to that community, and what does that mean for the community?

So to try to address this, we have put together The EMRG Program, Electronic Medical Records and Genomics, which is designed to tie together several of the U.S.-based biobanks and essentially to try address some of these key issues and move the field forward. So with that I really am done.

(Applause)

Joan Scott: Thank you. As Teri mentioned, and Kathy, a number of countries have national efforts, biobanks that are designed to be representative of the general population. And although there are many individual cohorts that are underway in this country, what we don't have is a nationally representative sample.

So one of the questions that has been raised by NIH and other federal agencies is: Should there be a national database of genomic and clinical information that's representative of the wider U.S. population, and what would people think about that?

So over the last two years, the Genetics and Public Policy Center has been involved in a public consultation project during which we sought to understand the public's attitudes about potentially having a national biobank.

In other words, should you build it; and if you do, will they come? And that's the information that I'm going to share with you today. The design that we were asking people about was a proposed design that has been floated out by NIH where up to 500,000 individuals that are representative of the wider U.S. population would be recruited, including entire households. And individuals or households that agreed to participate, those individuals would go to a study center near them where they would spend about a half a day getting a physical exam, providing information about their medical history, lifestyle information, environmental information, et cetera.

And then provide a biological sample for genetic analysis. That biological sample, blood sample, and that medical information would have their personal identifying information stripped to be sent then to this national repository.

A local key would be kept because participants would get the results back of that initial examination. But the information that was sent to the database would have the identifying information removed. That genetics study, then, and that clinical information, would sit as a resource that would be available for researchers to apply to to study common, complex diseases like heart disease, diabetes, arthritis, et cetera, et cetera.

And that information, then, could come back to either the database or medical literature at large. So that's the basic study design that we were interested in understanding what people thought about, would they be willing to participate, what concerns they might have about it, et cetera.

So over the two years we did 15 focus groups in five cities around the United States with a wide variety of individuals to understand what was the range of the perspectives around this issue. And then we tested those themes and issues in a large U.S.-based population survey of 4,659 individuals. We also did some interviews with community leaders and town halls, but I'm not going to talk about that data today.

What I'm really going to do is talk about information, interesting information that we heard from people in both the focus groups and the surveys about their attitudes around that. And specifically I want to touch upon a couple of issues. One is what would people think about developing such a resource, and would they be willing to participate?

Secondly, if they were participants, what would be important to them? Specifically around the desire to get information back and expectations around privacy.

So, first of all, people's support. We were very interested to hear and gratified to hear that people in the focus groups really understood -- even if they didn't talk about it in a sophisticated way -- they had a pretty sophisticated understanding that complex diseases like heart disease, obesity, et cetera, are due to both genetic contributors as well as environment.

People look at their own family history and sort of understand that. And they understand that in order to really tease apart the variety of genetic factors and environmental factors that contribute to these diseases, you need lots of people who are reflective of the wider population.

And when we in fact tested that in the survey, we found that that was true. Eighty-four percent of individuals in the survey thought that a national database or biobank for genomic research should be developed.

And, importantly, that support cut across all demographic groups. In addition, when we asked would you participate, 60 percent of individuals in the survey said that they would. And, again, importantly that cut across all demographic groups, and a majority of all demographic groups said they would participate.

When you talk to people about why they think developing such a biorepository or biobank is important and participating is important, here's some things we heard people say that we tested out in the survey.

Eighty-nine percent said that the study could lead to improved treatment. Eighty-four percent thought the study was important to do or said they were curious about the influence of genes, environment, and habits on health. And 81 percent said that participating would make me feel like I was contributing to society and it could benefit my family.

But when you ask individuals, so would you participate, we heard, yes, a lot of support. But the motivators and why people would support varied. In the focus groups, there were a number of individuals who would be willing to do it simply for the greater good. They had no expectations of anything back in return.

So as this person said in one focus group, it's pretty minor. If it's for the greater good, it's my way of contributing to everyone. There was a small number who said, no way, I would not participate. And usually that was around issues of trust. The fact that this was going to be a government-sponsored project. They were concerned about how it might be used, et cetera.

But a lot of people said yes, I would participate, but it really depended on some of the particulars and the devil was in the details. What we heard from people was that one important motivating factor that would influence their willingness to participate was, what they were going to be asked to do.

We heard things like, well, it sounds like a pain in the butt, it just takes up a lot of time. Or, it's very simple. I would do it.

We also heard people talk about what they might get back in return. Everything from a token payment as reimbursement for time to -- a little tongue in cheek, but maybe not the completely -- a nice little cruise. But what we also heard a lot was people interested in, can I get information back about myself? Not just the initial physical exam, but research that's done using my sample. I would want to know everything you find out about me. I would not want to do a study like this without getting results.

So we tested out some of these factors that we thought were important motivators for people to participate. And in the end of the survey we divided people randomly up into eight different study designs, which we varied depending on whether or not they were asked just to do a one-time physical exam or asked to do a lot more. Whether or not they were paid just \$50 as opposed to getting paid \$200, or whether or not they would get research results back or not.

And not too surprisingly, we saw that the willingness to participate ranged all the way from 73 percent for those individuals who were asked to do the least but got the highest payment and got research results back, down to 51 percent for those individuals who were asked to do the most, got the least amount of reimbursement and did not get any research results back.

In fact, getting research results returned was the most motivating of all of those three factors that we looked at, with reimbursement followed behind that. When we asked people, would you be willing to participate if you were not able to get research results back, 75 percent said they were less likely to do so.

People were interested in all different kinds of information. Ninety-six percent said that they would want to know about conditions that could be prevented or treated, but also 91 percent said that they would be interested in knowing about health risks even if there was nothing that they could do about it.

So people were interested about getting this information. Very few people resonated with not wanting to get research results back because it might worry them. Only 17 percent said that. It would be too much information or they were not interested, only eight and seven percent thought that.

Although people were really interested about getting stuff back about them, who else might have access to results was a concern. And so the confidentiality and the privacy around that information was important to them. We heard various things in the focus groups that people are pessimistic about the ability to keep databases secure. So we heard things like – there have been so many instances of laptop computers being left somewhere with information about people, and there's no way in the world that you can keep a secret.

But, more specifically, people were concerned about access from insurance companies. If they share this information, say, with health insurance companies, could we get denied health insurance. I don't think people want their employers to know because if they're paying for your health insurance and there's something wrong with you, that means you're a long-term risk and you're expensive, and would the police be able to get access if they knew that this man had murdered someone, for example.

As a matter of fact, on the survey, of all the list of concerns people would have as potential participants, protecting privacy was what they were most concerned about.

Specifically, 93 percent thought that if they were participants it would be important that it would be illegal for insurers and employers to get their information.

Now, since we did this study, the Genetic Information Nondiscrimination Act, or GINA, was signed into law in May of this year, and that prevents insurers and employers from requesting or accessing genetic information.

Those regulations will go into effect November 21, 2009. For information about how that process is going, you're welcome to go to our website. The passage of GINA may address some of those concerns that we were hearing from people about who could have access to the information.

But what about other individuals? Eighty-four percent thought that it should be illegal for law enforcement to get access to it. Imagine a scenario where there's a series of gruesome murders in Podunk, U.S.A., and there's DNA evidence at the crime scene, and Podunk U has a biobank in collaboration with a local health care system. Would the police want to get access to that biobank? Would courts grant a subpoena? What would the researcher and the university do? And would individuals who participated in the research and have samples in that biobank potentially be at risk for having participated?

Now, there are some potential protections for such individuals. Certificates of confidentiality are issued by the secretary, and that protects researchers from having to provide identifying information in the case of a civil or criminal or administrative case.

However, even though certificates of confidentiality are available for all research, even if it's not federally funded, the researcher has to ask for it. Not all researchers are, and even though the researcher's not compelled to release identifying information, they can if they choose.

So the case could be made to them, and they would say yes, in this particular instance with all these gruesome murders, I'm going to allow you access to this database.

In summary, then, there's strong support, we found strong support for a national database to study complex disease and a high willingness to participate. But potential participants are interested in

obtaining information about research that is done using their sample. And it comes with some expectations that there will be privacy protections in place.

So we might want to rethink about the nature of the researcher participant relationship. The human subject regulations are decades old and maybe there are some things that need to be relooked at in this era of genomic biobanks. We need to be attentive to the implementation of GINA as it occurs. Maybe we should think about strengthening certificates of confidentiality and making them obligatory. But most importantly, I think we need to make sure we're talking to the potential participants in these kinds of studies, because transparency will go a long way in building trust in this kind of research enterprise.

And with that I will end and thank the funders for this work, NHGRI.
(Applause)

Isaac Kohane: Just to warm you up, you're not ill yet, Mr. Blendel, but you've got potential... from the Promise of Personalized Medicine.

My name is Isaac Kohane. Most people call me Zak. And I'm here to talk to you about how we can use our already very expensive healthcare system to use both the informational and biological byproducts of the healthcare system for discovery research in the genomic era, without having to create an expensive parallel infrastructure of large cohort studies, maybe, as an alternative.

I'm going to be intentionally, as is my style, provocative. But it is to provide you with an alternative. Why do we need large numbers of patients? Well, to remind you all, there are rare diseases, there are common diseases. There are gene variants, which are highly deterministic, highly penetrant. And there are gene variants which are weakly effective, not predictive. You might have a variant that protects against Type II diabetes, but if you go to McDonald's there goes your protective effect, because it's dominated by environment.

There are hundreds of common diseases, but there are also tens of thousands of rare diseases. Although tens of thousands of diseases are rare, in total they may well be well over a million patients, where the common diseases are on the order of hundreds of millions of patients.

Lots of companies have done very, very well by focusing on these rare diseases, nonetheless. The diseases that we know and love, or hate, are part of the spectrum. And you need large numbers in either instance, whether it's looking for weak effects, where you need a statistically significant sample. For a large enough sample to be able to quantify that weak effect, you need large numbers of patients.

Or if you're looking at rare diseases and rare variants, you need enough patients to find those rare variants. Either way, you need large numbers of patients. In that I'm inspired by the world's first low-hanging fruit pickers. By your laughter, I take it that although we're not quite in the Bible Belt, you understand the allusion here. These are the first low hanging fruit pickers, and they're picking from the tree of knowledge. And to make sure you're awake, what did they first recognize when they picked from the tree of knowledge. They were naked. What great educations you all have.

So I will end by actually showing how this kind of discovery using the healthcare enterprise can actually make us understand how naked our healthcare system is.

Let me tell you about some of the tasks that we're doing at Harvard as part of our National Center for Biomedical Computing, for which I'm the principal investigator.

For example, can we predict, using genetic variants and clinical factors, who with asthma, after treatment with state-of-the-art glucocorticoid Prednizone therapy, will bounce back to the emergency room? Even if this would not lead to a new drug targeting those gene variants, it would be good to know because we would actually treat those patients differently and potentially save quite a bit of money.

And here's the problem: Genetics is cheap. There are a number of companies who seem to be making a business plan out of selling lots and lots of sequence for zero dollars or close to it. But the clinical factors are hard to come by. Description of the patients. Billing codes are too coarse-grained. Billing codes are biased for income. So what do you have left?

You have the narrative text inside the healthcare records that I just heard Barack Obama is going to be investing \$40 billion in. So what does it look like to actually go for a phenotype, for a characterization of a patient in the electronic health record? Here are some examples, just for those asthmatics. What's the smoking status? This is smoker, nonsmoker, never had tobacco. You say, Zak, what's so hard about this. 50 packs a year. Tobacco three weeks ago. Is that a smoker or a nonsmoker? Unclear smoking. Do you know what that is? Only if you're a slovenly resident as I was will you understand that at 2:00 in the morning, I don't remember what the patient said.

What about the lactobacillus? Isn't that a poor programming job? It's the same programmer who has to find it in tob/alcohol. So it is complex to actually find all these phenotypic characterizations in the natural language processed medical record, but it is doable.

So at the end of the day, at Partners Healthcare System we're able to identify 96,000 patients with asthma in a matter of days, identified out of the two and a half million Partners Healthcare System patients. Stratified by severity, pharmaco-responsiveness and exposures, and now with cases and controls, those who bounce back to the emergency room several times and those who don't.

We have three methods of tissue acquisition, of which I'll only describe one. So the context is this: The Secretary of Health and Human Services had a committee looking at genetics, and they said, we actually need to do a study with a million patients but we think it's going to cost about \$3 billion, which I used to think was a lot of money until the recent fiscal disaster.

But probably still for biomedical research it will be a lot of money. So the three prongs of high-throughput instrumentation enterprise includes something that, thank goodness, the biotech companies have done for us: Free or near-free genotyping, genetic sequencing.

The phenotyping, although it costs literally 20 bucks per medical record read, if you use the natural language processing techniques that I just described, is actually very high throughput, very cheap. But what about the samples? Typically NIH or a drug company gets charged upward of \$1,000 to \$3,000 per sample to get the biomaterials on a patient.

And I was despairing, as the principal investigator of this national center, how we would actually do this. Fortunately, as always, it's better to be lucky than smart. And my luck came in the form of a woman named Lynn Bry, a pathologist at Brigham Women's System, who had developed a system called Crimson, and Crimson works like this.

Let's say we've extracted, as we did, the 96,000 patients from the healthcare record who have asthma. We generate a unique code, a one-way hat, a code that you could not tell who the patient is, based on those 96,000 patients.

And then every time that patient comes back to the healthcare system, whether for a preop check, for a cholesterol check, or for routine care, and blood is drawn, when the sample is about to be discarded, after a 24-hour period or 48-hour period during which these samples are held onto for quality control purposes, we generate the same code, the same unique code, again irreversibly not traceable to the patient, and see if it matches any one of our 96,000 patients.

So if any of those 96,000 patients come back, we have anonymous sample linked to their anonymous phenotypic characterization.

And so we did a study in asthma, just as I was describing, and here is our projection of what our recruitment rates would look like over a ten-month period of Caucasian Americans and African Americans in Boston with high and low utilization of asthma services.

We estimated about 2500. It turns out that when we actually ran the study, after a slow start, we were actually able to get well beyond the projected number of samples, upward of 3,000 samples in the same period.

Now we're doing better and better. These are on a graph showing 50 weeks and hundreds of samples. We think we can do for common diseases such as major depression, asthma, and rheumatoid arthritis, 500 samples per week on average at a cost of \$8 per sample.

And so what does this mean? It means that the current cost for a 20,000-patient study is probably around \$25 million. Without any breakthroughs in genomics, we think we can get it done for about \$2 million.

That same study, which the Secretary of Health and Human Services thinks can be done for \$3 billion, we think can probably be done today for about \$1.2 billion. But if we use the system we just described, for \$160 million. This is in fact a disruptive technology. Two hundred magnitude faster sample acquisition and one order of magnitude faster -- cheaper. But let me switch gears. That's one version. High throughput, and it's only a one-time encounter with the healthcare system. You can't go back to the patient.

I've been working about 15 years on personally controlled healthcare records. There was a recent *New England Journal* article about these personal healthcare record systems, one we're responsible for having built, and our architecture informed all three major efforts by Microsoft, Google, and others.

I proposed, based on personal health records, a completely novel system called Informed Cohort that we published in *Science* in 2007. And this is how it works: It says that in a study, a patient gives you a full history, as they do today. They give you their electronic history from the healthcare system, or give you access to it, as they do today.

They give you a blood sample to do a genome-wide scan, to measure a million SNPs or half a million SNPs on a standard chip, as they do today. You then take that data and you anonymize, as we do today,

and put it into an anonymized cohort database. Then you have an analyst studying the cohort. Now, what happens if we actually find something important?

The reason I wrote this paper is because in my study of autism we found two patients who had some terrifyingly bad disease. I could not get back to them about it in an obvious way because of the mutual ignorance pact that doctors and researchers enter into in most of these blinded anonymous studies.

So what if you find something important, whether to the benefit of the patient or risk to the patient, a new drug or potential risk? What we do is: After going through the IRB, after going through a new committee called the informed cohort oversight board, we actually broadcast over the Internet in encrypted fashion a message that will uniquely arrive to the patient who now has -- and here's where it's different -- a personal copy of all that data that they disclosed to the anonymous database. They have their personal health record with all their data.

For those who are old enough to remember UHF television, you can either have your television off or you can have it on, in which case you'll receive messages. You can have it tuned to autism channel, to the cancer channel, or all channels, giving patients full autonomy about what kind of messages they get.

And, furthermore, we don't know who the patients are who we're messaging because we're looking at an anonymous database. But because we have their genomic signature encrypted, the personal health record will know that that message is uniquely destined for that patient and therefore, if the patient has selected to be reachable and on those topics, they can then be reached.

And so we actually have funding at the Children's Hospital in Boston for running this project, and we're doing the pilot this spring.

Last three slides, in fast order, I promised you nakedness of the healthcare system when you pick low-hanging fruit. Here's some nakedness. Here's Partners Healthcare System in 1997 to 2006.

What am I about to show you? A graph. A graph of what? Of a big peak. What's that a peak of? It's a peak of heart attacks. What kind of peak is that? It's an 18 percent increase in heart attacks.

Do we know this is happening? No. What was happening here? The arrow. Vioxx in. Vioxx off. Right under our nose -- and this was happening all around the country and we did not know it.

If we bothered looking at our healthcare data, if we could have, if we instrumented our healthcare enterprises, as we can do today, we would understand that nakedness. Right now we're naked because we're not doing this.

To give my co-PI, John Glaser, Partners Healthcare System, immense credit, rather than trying to find a deep pit to bury me in when I disclosed this, he actually said: How do we put in place the data safety monitoring board and governance procedures to actually interpret this data correctly.

By the way, we're able to do this early for a number of drugs. I won't tell you the other drugs so we don't short their stock. But here's a bit of a sad part of this same Vioxx thing. As a number of prescriptions went up in our system, the age of heart attack kept dropping.

This is not hard stuff to do if you do it well, and this is the kind of opportunity we have to understand the nakedness. And the nakedness is in many dimensions: genomic, pharmaceutical, and all the combinations within environment.

In summary, the low-hanging fruit have ripened through the application of bioinformatics information technology. Healthcare data can be the gateway to discovery science, including the results of the genomic revolution.

There are many possible selections in the space of autonomy, efficiency, and communications. Society must and will lead and all we can do as scientists is to provide the range of solutions.

And with that, I thank you very much.
(Applause)

Kathy Hudson: Thank you. Before we open up for your questions and comments, I have a couple of questions of my own. Zak, when you're allowing people to tune in in your model, into information that's being collected about them, that is unfiltered information? It's all information? Or it's edited? What information is that?

Isaac Kohane: So contrary to my first example, this is a very expensive system, because we actually -- now speaking -- I occasionally speak as a patient, but now I speak as a physician, we don't want to do harm. There's much in the genome that's not known, and a lot that's published ends up being contradicted.

So the short answer to your question is no. We use the IRB to decide what discoveries can be communicated to the patient because the burden of evidence is that it's actionable, and not only actionable, but actionable in a reasonable interval. For instance, if you find that a child that was just born has a high risk of Alzheimer's, it's not clear that the IRB is going to want you to disclose that.

Furthermore, we created this new institution called the ICOB, which, interestingly, I see CORIEL has mentioned that in the literature of the ICOB, informed cohort oversight board.

We have a communication specialist to say, what's the most effective way to send a message without unnecessarily creating fear and confusion? It might be, for most messages it might just be, we have something interesting to share with you at Children's Hospital. I don't know. I'm not a communications specialist. So the short answer is highly filtered, only clinically reasonable communications that you'd want in a server way as a clinician who wants to do no harm with wanting to communicate.

So maybe you could argue paternalistic, I would argue safe.

Kathy Hudson: In terms of the biobanks out there, CORIEL is one that's providing information back from the research study. Joan, are there others providing information back, or is the norm that the research information may be toxic and harmful and so we're going to keep it quiet?

Joan Scott: I don't know the rationale is because it's toxic or harmful. I think there are expense and infrastructure needs, and being able to responsibly provide information back to people will probably require a great deal of sitting down face to face and talking about what that means.

So some of it, I think, is a care issue around making sure that we're not sending back things that we don't know what they mean. But it's also that the predominant researcher paradigm under which we've been operating for a long time in that research results do not go back to participants, and you go into research with the anticipation that you're not going to get back anything directly.

Isaac Kohane: I do think that the results of your survey were fascinating, that the most informative factor for whether subjects would want to participate would be whether they would get information back.

I think ultimately that will be what will make this kind of paradigm succeed, because self-interested researchers will find they get the best cohorts, the most lasting cohorts. In fact, if you look at the large cohorts, like the Framingham Heart Study and other health studies, although they don't get back personalized information, they get back from the investigators a whole bunch of information so they really feel like they're participants.

Teri Manolio: In Framingham and similar studies we actually do give information back. One of the challenges is when is information actionable, when is it appropriate to provide it back.

But once research and the Framingham and certain studies showed high blood pressure is bad for you, you can do something about it, there were guidelines that that information was specifically fed back to somebody and they are actually set levels of, if it's too high you go immediately to the emergency room, et cetera, et cetera.

The challenge in genetics has been that it's kind of descended through the Mendelian era where finding a variant was, if not a death sentence, at least in many cases something would develop that you had no control over and it could be a really dreadful disease -- Huntington's is a prime example of this.

And after some efforts at trying to feed that information back to people, it became fairly clear that people didn't want it back for the most part and that it could actually be dangerous to provide it back to them.

Unfortunately, that's kind of led into the complex disease realm where you're talking about a risk factor that is actually less strong than knowing that your mother had a heart attack or knowing that your dad had diabetes. And I think we need to move the discussion along and the study will help do that.

Kathy Hudson: Let me ask another question, Teri. Zak talked about the cost to do the study that was recommended by the Secretary's Advisory Committee on Genetics, Health, and Society, which in fact was your study, and he quotes a price tag of \$160 million. Do you buy it?

Teri Manolio: We need to talk.

Isaac Kohane: Do I have a bridge to sell you?

Teri Manolio: We would be delighted if that were the case.

Kathy Hudson: With that, I'll open it up to you guys. If you'd raise your hand with a question or comment, a microphone will magically appear in your hand. Over here and then over there.

Can you please say your name and where you're from.

Question: Adam Clark with the Lance Armstrong Foundation. Very good presentations by all of you up there. We've been working with the NCI on adolescent and young adult biobanks, trying to build one. And an issue has come up about consent and re-consent.

And I was wondering, has this been addressed? Are you looking at this issue? Have you talked with parents about this or with family members? Any of you?

Kathy Hudson: In our study we asked people specifically what their expectations were for consent and what their expectations were for the nature of the relationship with the research project. So maybe Joan can talk about that and then Zak.

Joan Scott: It was interesting. People were almost split between those who wanted to not be bothered every time -- they wanted to get a consent up front to do whatever research projects you think is appropriate. The other half, though, wanted to be informed or re-consented for every potential study that their sample might be used for. So people were very divided on that issue.

Adam Clark: I think the issue we found was whether or not you're obliged to re-consent somebody when they go from --

Isaac Kohane: The answer is yes. Working at Children's Hospital, one of my colleagues, Ken Mandl, just published a paper in the *Journal of the American Medical Informatics Association* addressing just this issue, because it's actually three different consents. One is the consent that is essentially assent the parents give on behalf of the child. Then there's the major minors, the kids who should be involved in decision making. Then there's when they achieve majority.

And legally those re-consent processes have to occur, for better or for worse. Let me give vivid examples that really occur all the time. The patient is born genetically XY, looks like a female. And the parents decide, as they often do, not to tell the subject that they are an XY male but they in fact look like a female, are treated like a female, grow up as a female.

But then they discover -- and this has happened in the past -- they're in fact XY. They want to know. In fact, the old standard of care of pure paternalism has gone away. You actually have to re-consent. If they want to know, the burden is quite substantial. So it's an important issue. There are autonomy issues, but I think it's a legal issue you are asking, and I think we're legally obliged to actually re-consent.

Kathy Hudson: John, did you have a question?

Question: John Wilkerson, *FDA Week*. Where do biobanks fit into the comparative effectiveness scheme that's being proposed on Capitol Hill as a major piece of healthcare reform? And has anyone made sure that Baucus's staff is aware of the importance of biobanks?

Teri Manolio: I'm sorry, I'm not familiar with how they might figure in that. It would certainly be a reasonable place to do that kind of research, especially if they're linked very closely to the medical care system.

If they come from a hospital system where they're continuing to receive care, you could actually do work where, if you find a variant, you have the medical record, sort of ping the doctor say there's XYZ variant,

you need to follow this person up for Q,R, and S. I'm not aware of active outreach to biorepositories at present. I don't know if you are either.

Isaac Kohane: I don't think it's on the top part of the radar at this point.

Question: I know *FDA Week* might help that.

Question: This is Eric Bollinger. I'm a concerned citizen. In terms of the electronic health record, how mature or immature is the effort to standardize it? And do you see a centralized solution emerging or will it continue to be decentralized with the market determining the winner?

Isaac Kohane: You may not have intended it that way, but it's a very politically charged question, because, first of all, it's very topical.

Barack Obama has committed to a \$40 billion investment in electronic health record infrastructure. It's likely, though, in my very personal opinion, that there will not be a national central data repository, that instead there will be renewed effort to make the existing distributed databases in plain English, the records of patients in multiple hospitals integratable, if you have the right study design and the right IRB oversight to do it.

I don't think there will be a centralized solution but I think there will be increased emphasis on interoperability and sharing. There have been some efforts that have been successful in the last four years, but I think stepping up, making that effort more visible, will be quite important.

Question: Deborah Runkle with AAAS. Secretary Leavitt has been very interested in personalized medicine. And I'm wondering if anybody has a feel for the incoming secretary, presumably Senator Daschle, and whether his interests will be at the same level, greater, or less?

Kathy Hudson: I think we don't yet know. Secretary Leavitt did have a personalized medicine initiative. And he had many elements of that initiative and there were many, many committees that were formed and studied those issues. At the end of the day we didn't make dramatic progress, I think, in terms of actions or programs or policies that changed as a result of that initiative. That's my own view.

The incoming Secretary has some direct experience with genetics. Prominently, when he was leading the Senate for that very short period of time, he was responsible for moving the Genetic Information Nondiscrimination Act along. And his first press conference was, in fact, on the Genetic Information Nondiscrimination Act. So he knows the underlying science and is also concerned about the policy uses of that information.

Perhaps more directly relevant is that President-elect Obama, when he was in the Senate, had a bill on personalized medicine that had sort of two components to it. One was a component looking at the appropriate oversight of genetic tests as they move into the marketplace, and then the other part of that was actually focused specifically on biobanks and standardizing and making those systems work better.

So it will be interesting to see if he takes those interests with him as he moves into the White House. Does anybody else have comments on that? So we'll see.

Deborah Runkle: In other words, you don't know whether or not?

Kathy Hudson: No. But I have a microphone.

Teri Manolio: I would suggest reading his book on healthcare, and you can --

Kathy Hudson: Daschle's book.

Teri Manolio: Certainly there are opportunities to make the healthcare system more efficient. If one truly comes up with well-predictive variants. Whether they're predictive or not is still something that has to be proven.

Joan Scott: Zak, you had mentioned that in the model that you were using in order to give information back, that that goes through an IRB. How knowledgeable do you think IRBs are in this area to help make these sorts of decisions about when is something that is important enough or relevant enough that we should send it back?

Isaac Kohane: That's a very, very well-posed question. As few genetic counselors as we have in this country, we have even fewer IRBs able to deal with this kind of information. When we first posed this as a task, as a challenge for our IRB locally, they felt a little challenged.

It's just a lot of work. And so I think that it's both a lot of work and the knowledge base is relatively scarce. But I think this speaks to a larger deficiency in our medical workforce in terms of genetics.

Just to keep you entertained, there was a study done both in the Netherlands and the United States of primary care practitioners. What percentage of them do you think had ordered genetic tests for cancer screening in the prior year? Any idea? Price is right rules. Come on. You, you asked questions.

Question: No idea.

Isaac Kohane: Anybody want to venture a number?

Question: One percent.

Isaac Kohane: One percent is the modal number. Someone said 20 percent. That's actually closer to the truth. Most doctors will say one percent. The answer is a shocking 20 percent. Then when you ask yourself the question, what was the highest --

Teri Manolio: What did they order?

Isaac Kohane: Mostly breast cancer.

Teri Manolio: Colon cancer.

Isaac Kohane: Asymptomatic cancer screening. And what was the most predictive characteristic of that encounter?

Kathy Hudson: Receiving marketing materials.

Isaac Kohane: You are an appropriate cynic. There's a lot of truth to that. It's not that. It's not the age of the doctor. It's the patient asking for the test. It was the patient asking.

So here we have a healthcare system where the doctors don't understand the test. The patients see a webpage. They ask for a test to be done. The doctor doesn't know how to interpret the test unless they have a wonderful genetic counselor working with them. We're in a system where the healthcare system broadly, not just the IRB, is just not well positioned. We're doing this as a research pilot and we're doing it using, I would say, substantial resources.

Joan Scott: So what, if any, efforts are being made for those IRBs that are starting to have some experience like yours in this? I mean, sharing that information with other IRBs? Is there any mechanism by which they can do that?

Isaac Kohane: I would think that's a great project for your institute.

Kathy Hudson: I think the other complication here is if you look at large multi-center studies, you're going to have IRBs at all of those places, and you're going to have varying understandings of the underlying science and varying values that are going to be applied in terms of how protective versus not protective are you about the patient population.

So I think that's going to be an additional complication. When the IRB system was created it was sort of in deference to local values, and what we're seeing is just random heterogeneity, rather than a real reflection of differences in local values between what's happening at Children's versus what the IRB across the river thinks. So we'll see.

Question: Gail Javitt, Genetics and Public Policy Center. I want to go back to the question that was raised before about re-consent of minors and the answer that one is legally obliged to re-consent, which brings up a bigger issue: What's the legal framework, if any, that's governing the obligations of researchers to participants in biobanks? Is it under the common rule? Has the common rule thought through all these issues, or are the biobanks making these determinations for themselves in the absence of any guidelines?

Kathy Hudson: You're probably the person best -equipped to answer this question. I'll call on Gail Javitt. She's the attorney.

Isaac Kohane: I think it's a moving target. Let me tell you why it's a moving target. It's one thing to have measured one thing as part of a question, but if you're measuring everything, your whole genome, I think it's really -- I can't remember who wrote a very nice article in *Nature* about the blurred boundaries between research and clinical care when you start measuring everything.

It was a classic case where I think it was *in utero* testing for Tay Sachs disease where patients were informed of that, but because the way the consent was written, the way the study was written, they were not told whether or not they had the Trisomy 21, Down syndrome, and three kids were born with Down syndrome and the parents sued because that was knowable. So I don't know if we know what the obligations are for the biobank.

For instance, if you see on a SNP chip, which is basically a point-to list view of your genome, looking to the points, you see some large translocation that is known to be badness, whether developmental in utero

or for cancer. What is our obligation? Should we be scanning for it? Because I can tell you directly that most facilities don't have in place the computational/knowledge base wherewithal to actually be looking at all these chips to see if this is happening. They're asking narrow questions. But your question is an excellent one.

And my answer is, it's unfortunately a moving target. And unfortunately it may be lawyers who help push that boundary.

Kathy Hudson: Questions? There's one more back over here. Two more. So one there and one there and we'll draw to a close.

Question: I'm Matt Jones with *GenomeWeb*. And I follow NHGRI, and they and other parts of NIH, including NICHD and NIHS, are involved in starting large-scale cohort studies. There's the genes, environment, and health, and then there's the one that's going to do 100,000 children and follow them to the age of 20.

Can you talk about where we are right now in terms of what's going to happen with all this data and these specimens?

Teri Manolio: So the National Children's Study is the one that's the furthest along. And that is the 100,000 children that actually are being identified *in utero* and followed into birth and then beyond that. And that one is in the field.

I believe they have almost all of their sites selected and are gearing up to do recruitment. And they're funded at present. One would expect they would continue to be funded. The expectation, as I understand it, is that those data and samples will go into a repository managed by the NIH for a variety of Health and Human Services agencies and other agencies in the federal government that are collaborating in that study. And that they would be made available through appropriate controlled access procedures to any investigator who has a reasonable research question, et cetera.

The Genes and Environment Initiative is not a cohort study or a biorepository. It's a series of sort of preparatory steps in order to be able to do such work, primarily building on genome-wide association studies. So the chips that Zak was talking about, that really came into being and wide use in the research setting just within the past year or two, are being applied. And then the findings of those are being carried further into additional kinds of technologies because this is a technology, essentially.

The plan for the large-scale cohort that Zak and Kathy were referring to was to do genome-wide association genotyping and follow up the results of that to try to identify genetic and environmental determinants of disease.

As time goes by, sequencing becomes cheaper and cheaper and it may well be that we'll shift over, if we're finally able to do a large cohort study or any cohort study, into sequencing.

So we would have the complete genome sequence of an individual, which will again give us pause with respect to what to provide back to participants as well as how to interpret the data. But our expectation is that availability of those data would be very similar to the GEI program, where essentially any investigator who has an appropriate institution that agrees to maintain participant confidentiality and protect the data and use it within the informed consent would have access to it.

Question: Joe Devaney, Children's National Medical Center. My question is actually for Joan Scott. I was wondering, out of the 4900 people that you looked at, how much genetic knowledge did they have, and did that influence whether or not they're willing to participate in a survey?

Joan Scott: I should probably ask Dave, who is our main --

Dave Kaufman: Joe, we didn't really assess their knowledge, except to ask a question about what they believed. We gave them a list of factors of what they believed were possible causes of diabetes, and they could check genes, environment, on and on. And people all recognized that genes were important there. But we didn't really do any formal assessment. And that's a good question. It would be interesting to know.

We did ask if people had a genetic test before. Five percent did, and it wasn't related to support or willingness to participate.

Joan Scott: Although people's stated curiosity about genes and environment was correlated with their willingness to participate. So I don't know if curiosity and actual knowledge are related. But curiosity was an important factor.

Kathy Hudson: Please join me in thanking our speakers, and I'd like to thank all of you.
(Applause)