Public Health Perspective on Large Population Studies of Human Genetic Variation, the Environment, and Common Disease
Gilbert S. Omenn, M.D., Ph.D.

DR. TUCKSON: Gil Omenn, terrific to have you with us, and we're looking forward to your perspectives on the public health point of view on large population studies of human genetic variation, the environment, and common disease.

For our speakers, by the way, just so you'll know, there is a little timer that's sitting right beside Sarah, and if you want to gauge where you are, it's there with the usual yellow light.

DR. OMENN: Thank you very much, Reed.

It's a great pleasure to join you. This is a scenario in which I've been intensely interested for decades, at least 35 years in pharmacogenetics and ecogenetics. So the chance to at least share with you how I think about think this and how I think many people in the public health sciences and public health practice think about the opportunities to really make a difference as we expand our knowledge base from genetics and other fields is especially welcoming. Thank you for having me.

So here is a visual image which actually is a short-term vision, but we'll carry on for decades of work. As you've just heard from Dr. Goldstein, we already have the beginnings of an avalanche of genomic and genetic information, validated SNPs, the beginnings of a haplotype map for applications, candidate genes and alleles, and especially many candidate genes and alleles for particular disease risks.

The second bullet has been very much less addressed, and this is the improvement of our environmental and behavioral data sets and, most importantly, their linkage with genetic information. In fact, we have many proposed statutes and regulations that would make this impossible. I'll come back to that at the end.

The third is, of course, to carry out both of the first two items with well-established and, in the public mind and the legal mind, creditable privacy and confidentiality protections, both for genetic and non-genetic information. I'll come back to that also.

Finally, I think we can be quite confident that the technologies we have in hand and the concepts that are being developed will yield breakthrough tests, vaccines, drugs, behavior change schemes, and regulatory actions, all of which would be aimed at reducing health risks and treating patients cost-effectively in this country and globally.

You know, in medicine, we say we save one life at a time. The School of Public Health at Johns Hopkins has adopted this wonderful logo: "We save lives millions at a time." That's the public health perspective.

The world in which we live is well known to all of you here. We're very excited about the new biology. Most of us recognize that many of the developments in the biology have been made feasible, even conceivable, by new technologies.

You know, there's this notion you go from science to technology to application. Well, there's a huge feedback loop from technologies. This is reflected in gene expression microarray,

comparative genomics, proteomics, in which I'm working intensively these days, bioinformatics and computational biology, and on the medical side, and increasingly the community health service and public health preventive service's side, we talk about evidence-based medicine.

How many of you have heard that phrase, "evidence-based medicine"?

(Show of hands.)

DR. OMENN: Well, when we use it at a Rotary Club talk or someplace else, you can see the mouths open, the jaws drop, and finally somebody articulates the question if this is exciting and new, what have you folks been doing up until now? It's a little embarrassing. But we're doing better. We're trying hard and, of course, sometimes the hardest sell is with our own clinical colleagues.

The vision from all this is a kind of health care and community-based services that would be personal, predictive, and heavily preventive.

This takes people prepared to carry out such programs. The Institute of Medicine two or three years ago issued this report, in which they stated "With the arrival in which we will have the ability to understand gene/environmental interactions comes not only the era of genomic medicine, but of genomics-based public health. Understanding genomics, therefore, is essential for an effective public health workforce."

The CDC is particularly well represented here today, appropriately so. Here are our centers that CDC established several years ago already, including one we're proud to have at the University of Michigan, another which I was pleased to help get started at the University of Washington, and the third in North Carolina. They collaborate effectively. They have a website you can check. The mission is exactly the mission of this discussion.

Now, just so we're on the same wavelength and especially those who are likely to be aware of this meeting, and I'm not actually integrally involved, definitions do matter. There's something of a struggle over which is the broader term, "genetics" or "genomics." In quarters where I live and in some recent reports, we've tried to help the public and help ourselves understand genetics as the broader historical, broader scientific term of approaching genes and their roles in health and disease, physiology, and evolution, and genomics being the set of powerful new tools for molecular biology, biotechnology, and computational sciences that permit us, when we choose, to examine the entire complement of genes and their gene products altogether, although, as you've just heard, generalizing across all genes is a formidable task and we end up focusing pretty quickly.

These global analyses do permit us -- in fact, require us -- very usefully to go beyond what we sometimes speak of as "looking under the lamp-post," where we already know about a gene or a phenotype that we're most interested in or a desired effect from a drug and ignore the off-target actions of the same drug which lead to nasty complications and cost of the drug.

The same thing on the protein side. We can talk about individual proteins or proteins as a class. We can talk about proteomics, corresponding to genomics, looking globally at as many as possible of the very much larger number of proteins and protein forms that are coded for by those genes.

So we already had a good introduction to this subject about genomic information from the global analyses, the International HapMap Consortium, the direct associations of individual SNP alleles with various disease phenotypes, the very substantial database -- we heard it's now over 10 million -- and the haplotype structure work, which is really still emerging with a lot of clever efforts to use tagging SNPs and variable linkage disequilibrium, recombinant hot spots, and other details of haplotype structure.

Where can we get information about environmental variables to put together with the genetic information? Well, I'll give you a few examples, and you'll more from Dr. Manolio and others this morning.

The Centers for Disease Control National for Health Statistics has conducted for 40 years surveys of the American population and increasing numbers of laboratory analyses. Now, we're going to hear later and I will come to a slide about what is the set of categories called "environmental" or "non-genetic" in the U.K. Biobank, but here I want to focus particularly on chemical, microbial, and, say, environmental exposures complementary to behavioral traits, reproductive history, and others which you will hear more about from others.

The NHANES, as it's now called, is proud of major impacts. It's a major contributing factor in the removal of lead from gasoline, one of the public health triumphs of the last century, elaboration of pediatric growth charts, prevalence estimates for cholesterol, blood pressure, hepatitis C, and other important variables.

These are the environmental exposures that are actually assayed currently in the NHANES, and this is ongoing. So lead in a lead biomarker in sites, cadmium, mercury, arsenic, organic chemicals, acrylamide, which is a reproductive and neurotoxin, phthalates, metals, IgE antibody showing latex allergy, aromatic hydrocarbons, phytoestrogens, dioxins, and a whole bunch of usually serological markers of microbial exposures. Also, cotinine for smoking history or, if a non-smoker, environmental tobacco smoke exposure, and a whole lot of other phenotypes measured in the laboratory.

So this is a rich data resource. Over the years, the NHANES II, which concluded in the '80s and had 14,000 people. NHANES III, 34,000 people. I actually couldn't find in the very extensive website of NCHS the number for the current ongoing NHANES study. Muin Khoury told me that there will be about 6,000 or 7,000 so far who have DNA samples taken. I think that might be about a 10 percent sample of the total.

NIEHS is interested in environmental and genetic interactions. I recently have served on an Advisory Committee on Personalized Exposure Assessment. The approaches that we highlighted in our report, which will be out shortly in Environmental Health Perspectives, were the use of geographic information systems, and the example there is the NIEHS set of children's health studies, where they combined GIS and wireless devices to track exposures to pesticides to validate diary entries. These are diary entries not just of diet, but of activities and potential activities that would be tied to those exposures, including children who might be exposed as migrant worker families or children who would be exposed with concomitant information about pesticides in the house and garden, and they are developing spatial models for households at risk for lead poisoning and a variety of other exposures.

The second comes from the technology side of biosensors and nanoscale devices which will permit feasible measurement in the individual of exposures and relate then to actual bioburden measures of the sorts that NHANES does.

The third category is molecular signatures of exposure, early effect, and variation to susceptibility, which we call toxicogenomics. The conceptual strategy here of really building a program which would fit very nicely with what was just described and what's going to be described in the Biobank and some other large prospective studies may be applied in proper settings to retrospective or nested case-control studies as well, of course.

You have to be able to identify what your priority diseases are and the plausible or hypothesized environmental factors. This is non-trivial. In fact, we basically punted in this study for later work to be done on this.

Identify potential genetic determinants, pathways, and model systems for exploring the genetic/environmental interactions. Identify target study populations for feasible measurement. Define the genetic determinants of susceptibility. Conduct targeted exposure assessments. Identify and validate biomarkers. Then try to bring this all together with genetic/environmental interactions.

One thing that should be emphasized is that the era of fighting between whether things are nature or nurture, genetic or environmental, is behind us. We're now all thinking about contribution of genetic and non-genetic factors and specific ways they interact and even, I would say -- I cringed a little at the comment in the last talk that for Mendelian disorders, of course, we know exactly what the genotype/phenotype pattern is. It's a lot more direct than for multifactorial diseases, but it is also true that the variation can be quite stunning for single gene disorders, the most dramatic being reports over the last decade from Saudi Arabia and Jamaica of people with hemoglobin beta S homozygote status with no apparent phenotype, clinical phenotype, full biochemical phenotype, and many other examples.

Technologies and approaches. Some are listed here. I think I've already basically mentioned them.

This is natural process language to try to search the vast literature. There are some very good tools now becoming available for doing this in an automated way to us limited humans.

GIS I've mentioned.

Mapping and systems, and one of the questions I asked Muin was the extent to which the NHANES findings that sample all through the American population are actually being mapped as the EPA tries to do for other purposes to states, localities, neighborhoods, and maybe impute it all the way to individuals, and so forth.

This is one of the most important things for laboratory scientists, which is to link perspective sensors and molecular biomarkers in animals and in humans with in vitro and in vivo studies to try to make that between toxicology and epidemiology which has been needed for so long.

EPA. EPA, of course, regulates air, water, soil and, together with FDA, foods, food contaminants. The EPA has many measurement and modeling programs, of which this may be the most relevant for our purposes today. It's called the Multimedia Integrated Modeling System, MIMS. The primary application is to simulate ambient airborne substances in urban settings, and the spatial scales they are looking at range from 10 kilometers down to less than 1 kilometer, which gets to be interesting for imputation of individual exposures.
They are working on prototypes and successive generations of exposure modeling support tools, and this is both for air pollution and for homeland security. You can easily imagine that.

These tools bridge modeling gaps between two previously quite different approaches. One is the Eulerian chemical grid modeling and the other is the Gaussian plume dispersion models, which are prominent for water as well as air pollution. These models will capture temporal and spatial variability at ground-level concentrations of air toxics. Also, hazardous releases from stationary sites, and may reveal enough hot spots to be quite interesting in terms of human studies.

There is a sort of progression to make ambient measurements in the air wherever there's a monitoring station, and where those stations are placed, of course, is highly irregular and never been systematized around the country.

There are personal monitors. We're familiar with these in the workplace, of course, in industrial hygiene, but they're available for community sampling studies.

There is biomonitoring, as shown here for several examples. Of course, with biomonitoring in isolation, as with NHANES, or with maybe the studies that are going to be done under these genetic population studies, there's usually very little information about the source of the agent that's measured, and that needs to be thought about in advance.

Finally, there's the National Scale, sort of the summation of all this, and the CDC in 2003 already did have 116 environmental chemicals, including the ones I listed for you a moment ago.

Here, John, is my take from the Web and from a meeting I was at, a planning meeting in Dublin four years ago. I wasn't aware when I prepared my slides that we were going to have an expert talk about this from the people who are actually doing it, so I'll be very quick, but maybe it would be interesting to see the perspective of someone across the ocean about what we know about what's going on.

So this is a genetic databank to be developed from blood samples from half a million people. I understand that the studies will be based on proposals from researchers. The recruitment will be through general practices, many of them, in regional combines with a 10-year follow-up. The age at recruitment, 45 to 69, and there are expected to be substantial number of deaths over that period of time from common diseases, some of which would be of great interest here.

There will be a questionnaire for risks, lifestyle, diet, and there will be a blood sample taken. There's not been too much said yet about what the blood sample will be used for. Maybe we'll hear today.

Statistical power estimates. It's very important in planning studies, of course. They expect over 5,000 cases per year for diabetes mellitus, ischemic heart disease, myocardial infarction, colorectal cancer and breast cancer, and you can see here the projected relative risks and interaction ratios that they would be able to detect with these numbers and that power. I'm sure that should be .01. So 1 percent significance. Then at a lower incidence, there would rheumatoid arthritis, Parkinson's disease, hip fracture, ovarian cancer, bladder cancer, and others, with, again, power estimates.

They have a very high expectation that 40 to 50 percent of the patients in each practice would actually enroll. This would be astonishing in America. Maybe they can do it in the U.K.

Now, they've chosen for the blood sample EDTA plasma. It's a very interesting question always of what form of serum or anticoagulant to be used. In a separate big international collaboration I

lead about proteomics of plasma and serum, we have similarly given high grades to EDTA, but even higher to citrate plasma.

There will be nested case-control and cross-sectional studies, including a variety of family-based studies.

There have been some criticisms of the design, naturally. One is that even at half a million people, the cohort is much too small to analyze complex multifactorial diseases.

Heterogeneity within these disease diagnostic categories is extreme. When I was in Ireland, there was a big discussion about a proposal to actually enroll sib pairs, which would be particularly informative for genetic studies. I'm curious what the status of that is. I couldn't find any mention in the website.

The cohort age of 45 to 69, of course, is a late time to be gathering information about the crucial determinants of early stages of latent diseases, long-gestating diseases.

Of course, relying on medical records, while maybe they are better than here, is still a limitation. There is some comment that there might be an overemphasis on genetic factors because of the reliance on the medical record and because of the lack of much collection about other kinds of environmental factors, and there have been vigilant consumer and patients looking out for confidentiality and opposing any kind of genetic behavior studies, and some other concerns.

These are the exposure categories, as I understand it. You can see them all listed here, and no specific mention of environmental chemicals, which in this country would be top of the public's list.

Examples of the kinds of studies that can be undertaken you see here. All of them are interesting, yet they're of a subset of the variety that I've been indicating would be a broader environmental/genetic interaction.

Now, other large-scale studies are underway in various places, and in the Biobank site they mention the much-publicized studies in Iceland and less publicized in Estonia and under development in Canada. There's a big European collaborative study called EPIC, and there are others which Teri Manolio I guess has provided those of you who received the materials for this meeting.

Now, in this country, the most remarkable study of the last decade has been the Women's Health Initiative, with 160,000 women participating in both observational and randomized studies, and as you know, the outcomes have been front-page news most months.

Now, let me bring this into a little broader perspective from the public health view. This is about genetics and environment and how we share a lot of interests. We both aim to bring together the digital code of inherited information with the environmental cues, some people call them, from nutrition, metabolism, lifestyle behaviors, pharmaceuticals and nutriceuticals -- don't forget the nutriceuticals -- and these chemical, physical, and infectious exposures.

The broad way to think about this is a systems biology approach that looks at the inputs, the perturbations, and then genomic, epigenomic, transcriptomic, proteomic, metabolomic levels of integrating the molecular information.

Ecogenetics has been the focus of my talk here and I'm going to carry on for a few more minutes about environmental and occupational exposures and variations to susceptibility, but it can be looked at from the point of view of infectious diseases, chronic diseases, nutrition, unhealthful behaviors, and it means that we should include genetics prominently in protocols for health promotion and disease prevention, and these would include host/pathogen interactions as well as drug and vaccine development. I've already mentioned the training need.

Put all that together and there should be, in the next decade or two, really a golden age for public health sciences. We need these kinds of population-based disciplines in order to make sense of genetic variation. It would be a tragedy, in my view, if we had extensive genetic variation and really could not make the relationship to phenotypes or answer people's questions about what you could do with this information to reduce your health risks.

With regard to the chemical exposures specifically, there is a discipline of risk assessment, risk management, risk perception, and risk communication which has developed over the last 25 years. It's really all addressed at this question or this observation: scientists disagree.

This is extremely bewildering and disconcerting to a lot of people. In fact, in this current debate about faith-based ways of thinking and scientific ways of thinking, the characterization of scientific ways of thinking as all based on fact and certainty is a huge failure of our communication because we are typically most interested in what we don't know and what is uncertain and how we could learn more and make it useful.

There's a framework for this kind of thing with regard to regulatory decision-making in chemicals, and other factors, too, but especially for chemicals to identify whether there's potential for hazard with all of these methods, especially the ones I've been talking about, to characterize the risk -- very important word, characterize, not just to quantify, but to describe, have a useful narrative about the nature of the health effects observed, the phenotypes and how reversible they are, how serious they are -- related to potency, exposure analysis, which until recently was very under explored, and our point here, of course, variation susceptibility, and then to do something about it. Very often information, long before there's a regulatory action, has a powerful effect.

Toxicogenomics I mentioned. This is the signature program at the NIEHS, the National Toxicology Program. There's a framework which says we need to put any environmental scare or scientific finding into broader public health and maybe even ecological context, and then have an orderly process of developing an assessment of the risk, reasonable options, make decisions, actually make decisions and carry them out, and evaluate what we've accomplish if we do. All of this, from the very beginning, with active engagement, proactive engagement, of stakeholders -- very important -- as the genetics community has been doing around our issues.

Context means, in the environmental world, going beyond the statutory scheme we have of one chemical, one environmental medium, one health effect at a time. Think about the total public health status of children or of any other group.

Intense requires multiple molecular markers and especially a public health comprehensive view.

Context means medical source of the same agent, number of pathways of exposure, multiple risks from one agent or multiple agents causing the same effect, data, surveillance, interaction with the environment, and crucial issues about health disparities, environmental justice, social and cultural traditions, and differences in perception about risks and what should be done about them.

Finally, I want to point out some good work from an organization called Partnership for Prevention engaging with the states. Of course, CDC is very active with the states and other agencies. There's a lot of action at the state level. In fact, pending federal legislation on protecting people from insurance or employment discrimination for genetic diagnoses, some 38 states at least have passed their own patchwork of legislation.

Well, the aim for states is shown here. Monitor what's happening and to ensure that we have applications not just for treatment of people with specific diseases, but for health promotion and disease prevention.

These are the two key findings. The first we've already covered, that there's a lot of opportunity in this genomic era.

The second is a hot policy debate and it was the position of the Partnership for Prevention that genetics and genomics should be integrated into existing health, social, and environmental policies, rather than establishing stand-alone genetics programs. Maybe you don't all agree with this, but let me tell you why.

This is quotation from that report citing a very highly regarded report which I was not personally involved in at the State of Michigan from the Governor's Commission on Genetic Policy and Progress. "At a time when many state policies were based on exceptionalism" -- that means taking genetics out from the mainstream of medicine and public health -- "Michigan adopted an integration perspective and recommended that genetic issues be dealt with in the context of overall medical care values and principles."

"All health conditions have some degree of genetic basis. It's very hard to draw a line between what is genetic and what is not. Most common diseases that we're emphasizing here result from gene/environment interactions. So genetic advances are likely to extend and expand, certainly not supplant, current practices in medicine, public health, and environmental protection.

"Some genetic variations are associated with greater health risk than others. Covering this huge range with a one-size-fits-all policy is inappropriate.

"Decisions about genetic policy involve complex issues about ethics, costs, benefits, individual and societal interests. Medical care decisions should be linked with research, insurance, and broader public health policies.

"The intersection between genetics and public policy is both immediate and long-term, warranting close monitoring."

I added this line on the bottom, which is that in this era where in the clinic, where I will be all day tomorrow, we have to tell patients that it would be wise to make sure your insurance is complete and adequate before you have any tests done, and that prohibiting discrimination based on test results or genetic diagnosis is necessary.

The kinds of research we want to stimulate in populations and communities requires certain principles. Albert Johnson, a prominent bioethicist, observed in one of our seminars years ago in Seattle that while we had developed very widely accepted concepts and tools for ethics in medicine -- namely, the informed consent principle and the principle of autonomy of the individual participant -- that we had no corresponding highlighted principles for public health or community-based research.

So Jim Ledrefow and I and others developed and we published this scheme about engaging community partners early in the planning process, keeping them posted, seeking their input in the analysis and interpretation, building productive partnerships that last, and empowering people to propose studies.

There are sources of information shown here, and a final comment six years ago from Francis Collins that what we're engaged in collectively, mapping the human genetic terrain, may rank with the great expeditions.

It's clear that to get maximum value and meet our public responsibilities that we need to understand the progression from genes through proteins and some molecular and laboratory interests, and of course, clinical translation and, more broadly, to address the issue of this meeting, which is to link genetic variation with the many kinds of non-genetic variables.

Thank you very much.

DR. TUCKSON: Terrific. Thank you very much, Gil.

Again, any one particular question?

(No response.)

DR. TUCKSON: Thank you, Gil. We'll come back to you in just a bit.