Meeting Minutes
Department of Health and Human Services
National Institutes of Health
National Diabetes and Digestive and Kidney Diseases Advisory Council
May 12, 2010

I. CALL TO ORDER
Dr. Rodgers

Dr. Griffin P. Rodgers, Director, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) called to order the 183rd meeting of the National Diabetes and Digestive and Kidney Diseases Advisory Council at 8:30 a.m., Wednesday, May 12, 2010, in Building 31, C Wing, 6th Floor, Conference Room 10, NIH Campus, Bethesda, Maryland.

A. ATTENDANCE – COUNCIL MEMBERS PRESENT

Dr. David Altshuler
Dr. Nancy Andrews
Ms. LaVarne Burton
Dr. Charles Elson, III
Dr. Robert Flanigan
Dr. Christopher Glass
Dr. Gregory Gores
Ms. Jane Holt
Ms. Judy Hunt
Dr. Francine Kaufman
Dr. David Klurfeld
Dr. Brian Monahan
Dr. Mark Magnuson
Dr. William Mitch
Dr. Anil Rustgi
Dr. Anthony Schaeffer
Dr. John Sedor
Dr. Patrick Tso

Also Present:
Dr. Griffin Rodgers, Director, NIDDK
Dr. Gregory Germino, Deputy Director, NIDDK
Dr. Brent Stanfield, Executive Secretary, NIDDK Advisory Council

B. NIDDK STAFF AND GUESTS

Abankwah, Dora – NIDDK
Abraham, Kristin – NIDDK
Akolkar, Beena – NIDDK
Ameen, Vanessa – NIDDK
Appel, Michael – NIDDK
Arreaza-Rubin, Guillermo – NIDDK
Barnard, Michele – NIDDK
Beckley, Carey – NIDDK
Bethea, Gina – NIDDK
Bishop, Terry – NIDDK
Blondel, Olivier – NIDDK
Bloom-Davila, Maria – NIDDK
Calvo, Francisco – NIDDK
Carrington, Jill – NIDDK
Castle, Arthur – NIDDK
Chamberlain, Joan – NIDDK
Chianchiano, Dolph – Nat. Kid. Found.
Curtis, Leslie – NIDDK
Densmore, Christine – NIDDK
Doherty, Dee – NIDDK
Doo, Edward – NIDDK
Edwards, Michael – NIDDK

Eggerman, Thomas – NIDDK
Eggers, Paul – NIDDK
Ehrhardt, Britt – NIDDK
Evans, Mary – NIDDK
Farishian, Richard – NIDDK
Fonville, Olaf – NIDDK
Fradkin, Judith – NIDDK
Froyd, Erica – Lewis Burke Assoc.
Gallivan, Joanne – NIDDK
Gansheroff, Lisa – NIDDK
Garfield, Sanford – NIDDK
Garofalo, Robert – CSR
Goter-Robinson, Carol – NIDDK
Greene, Lucy – NIDDK
Grey, Michael – NIDDK
Guo, Xiaodu – NIDDK
Haft, Carol – NIDDK
Hamilton, Frank – NIDDK
Hanlon, Mary – NIDDK
Harris, Mary – NIDDK
Hilliard, Trude – NIDDK
Hoof, Eleanor – NIDDK
Hoofnagle, Jay – NIDDK
Horlick, Mary – NIDDK
Hunter, Christine – NIDDK
Hyde, James – NIDDK
James, Stephen – NIDDK
Jones, Teresa – NIDDK
Jones, David – NIDDK
Karp, Robert – NIDDK
Ketchum, Christian – NIDDK
Kim, Sooja – CSR
Kimmel, Paul – NIDDK
Klausing, Thomas – NIDDK
Kranzfelder, Kathy – NIDDK
Kuczmarski, Robert – NIDDK
Kusek, John – NIDDK
Laughlin, Maren – NIDDK
Le, Todd – NIDDK
Magra, Amy – NIDDK
Malik, Karl – NIDDK
Manouelian, Denise – NIDDK
Margolis, Ronald – NIDDK
May, Ken – NIDDK
McKeon, Catherine – NIDDK
Miles, Carolyn – NIDDK
Miller, David – NIDDK
Miller, Megan – NIDDK
Moxey-Mimms, Marva – NIDDK
Mullins, Christopher – NIDDK
Narva, Andrew – NIDDK
Newman, Eileen – NIDDK
Nicholson, Katherine – NIDDK
Ostell, James – NCBI
Patel, D.G. – NIDDK
Perry-Jones, Aretina – NIDDK
Pike, Robert – NIDDK
Pope, Sharon – NIDDK
Rada, Beth – XOMA
Rankin, Tracy – NIDDK
Rasooly, Rebekah – NIDDK
Roberts, Tibor – NIDDK
Robuck, Patricia – NIDDK
Rushing, Paul – NIDDK
Rys-Sikora, Krystyna – NIDDK
Sahai, Atul – NIDDK
Salomon, Karen – NIDDK
Sankaran, Lakshmanan – NIDDK
Sato, Sheryl– NIDDK
Savage, Peter – NIDDK
Sechi, Salvatore – NIDDK
Serrano, Jose – NIDDK
Smedberg, Paul – Am. Soc. of Nephrol.
Smith, Philip – NIDDK
Spain, Lisa – NIDDK
Star, Robert – NIDDK
Tatham, Thomas – NIDDK
Tinkler, Emily – NIDDK
Torrance, Rebecca – NIDDK
Vaunti, Patricia – Tech Team
Van Raaphorst, Rebecca – NIDDK
Wallace, Julie – NIDDK
Wellner, Robert – NIDDK
Woynarowska, Barbara – NIDDK
Wright, Daniel – NIDDK
Yanovski, Susan – NIDDK

C. ANNOUNCEMENTS
Dr. Rodgers

Elections to the National Academy of Sciences

Dr. Rodgers reported on the recent election to the National Academy of Sciences of one NIDDK intramural scientist and several NIDDK grantees.

Division of Intramural Research

- *Attila Szabo, Ph.D., Laboratory of Chemical Physics, Theoretical Biophysical Chemistry Section.* Attila Szabo, has conducted research at NIDDK for almost 30-years and is considered one of the world's leading theorists in the biophysical sciences. In addition to being elected to the National Academy of Sciences, he has been elected to the American Academy of Arts and Sciences, a prestigious institution composed of highly regarded scholars and accomplished individuals. A hallmark of Dr. Szabo's research is the study of the close relationship between theory and experiment. He has made important contributions in elucidating the dynamics of ligand binding to proteins and receptors on cell surfaces and the analysis of probes monitoring fluorescence depolarization and nuclear magnetic relaxations in both macromolecules and membranes. Dr. Szabo is best known for his model-free approach for the interpretation of nuclear magnetic resonance experiments. His pioneering paper on this subject has been cited more than two thousand times.

Division of Diabetes, Endocrinology and Metabolic Diseases (DEM)

- *Roger Cone, Ph.D., Professor and Chair of the Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine.* Before joining Vanderbilt University in 2008, Dr. Cone had spent much of his career at the Oregon Health and Science University. His research focuses on how the central nervous system regulates energy stores. He has specifically concentrated on the melanocortin system and its control of feeding and metabolism. Dr. Cone has been an NIDDK grantee for more than twenty years.

- *Robert Fletterick, Ph.D., Professor of Biochemistry, Department of Biochemistry and Biophysics, University of California, San Francisco, School of Medicine.* Dr. Fletterick has wide-ranging scientific interests. For example, his laboratory solved the first structure of a nuclear receptor bound to its hormone. His laboratory also solved the structure of the molecular motor kinesin found in nerve cells, showing that it is related to the well-known myosin motor protein of muscle. Presently, his laboratory is studying hormone receptors that regulate neurodevelopment, embryogenesis, steroid metabolism, prostate development, and cancer. Projects are also underway in regenerative medicine and stem cell science. Dr. Fletterick has been supported by NIDDK for different projects for more than thirty years. He also has substantial support from several other NIH Institutes (including the National Cancer

Institute, National Institute of General Medical Sciences, and National Institute of Arthritis and Musculoskeletal and Skin Diseases).

- *Richard Van Duyne, Ph.D., the Charles E. and Emma H. Morrison Professor of Chemistry, Department of Chemistry, Northwestern University.* Dr. Van Duyne is a long-time National Science Foundation grantee and is presently funded by NIDDK to work on a new *in vivo* glucose sensor technology.

Division of Digestive Diseases and Nutrition (DDN)

- *Ruslan Medzhitov, Professor of Immunobiology, Yale School of Medicine, and a Howard Hughes Medical Institute Researcher.* Dr. Medzhitov currently has an R01 grant from NIDDK investigating the role of commensal microorganism and Toll-like receptor interactions in the maintenance of intestinal homeostasis, tissue protection and repair, and the pathogenesis of inflammatory bowel disease.

Division of Kidney, Urologic and Hematologic Diseases (KUH)

- *Vann Bennett, M.D., Ph.D., a Howard Hughes Medical Institute Investigator, and James B. Duke Professor of Cell Biology, Departments of Cell Biology, Biochemistry, and Neurobiology, Duke University Medical Center.* Dr. Bennett has devoted much of his scientific career to the study of red blood cell membrane proteins. This work led him to discover "ankyrin"--the first known molecular connection between a membrane ion transporter and the cytoskeleton. He then went on to discover that ankyrin organizes specialized membrane domains in many cells in the body, ranging from excitable membranes in the heart and nervous system to epithelial cells, photoreceptors in the eye, and striated muscle. Dr. Bennett was a long-time NIDDK MERIT awardee before becoming an HHMI Investigator.

NIH Obesity Research Task Force

Dr. Rodgers informed the Council about activities of the NIH Obesity Research Task Force, which he co-chairs. The Task Force is developing a new Strategic Plan. Since the initial Strategic Plan was published in 2004, there has been considerable progress in obesity research--much of which has been supported by NIDDK and other NIH Institutes and Centers. The new Plan will reflect the current challenges and exciting scientific opportunities that have emerged. The draft encompasses research on discovering biological mechanisms regulating energy balance; understanding correlates, determinants, and consequences of obesity; designing and testing interventions; disseminating and implementing research; improving measurement tools, technology, and methods; and moving research findings into practice. To gather external input, the NIH sent out an initial working draft in April to scientists in obesity-related research fields--including some members of the NIDDK National Advisory Council--as well as to voluntary and professional health organizations. A revised draft will soon be posted on the NIH website for a public comment period. Dr. Rodgers recognized the contributions of NIDDK Program Director, Dr. Christine Hunter, as well as the work of many other

individuals who are providing support for this planning effort. It is expected that the final version of the Strategic Plan will be available in the fall of 2010.

II. CONSIDERATION OF SUMMARY MINUTES OF THE 182nd COUNCIL MEETING

Following a motion, the Council approved by voice vote the Summary Minutes of the 182nd Council meeting.

III. FUTURE COUNCIL DATES

Dr. Rodgers asked the Council to review the following upcoming meeting dates.

2010
September 22-23 (Wednesday and Thursday)

2011
February 16-17 (Wednesday and Thursday)
May 11-12 (Wednesday and Thursday)
September 7-8 (Wednesday and Thursday)

2012
February 15-16 (Wednesday and Thursday)
May 16-17 (Wednesday and Thursday)
September 12-13 (Wednesday and Thursday)

While most meetings are expected to be a single day (Wednesday), Dr. Rodgers asked Council members to reserve the following day (Thursday) to ensure flexibility in the case that a situation arises where a longer meeting is required.

Dr. Rodgers reminded Council members that NIDDK's 60th Anniversary celebrations will culminate on Tuesday, September 21, 2010 (the day before the next Council meeting), with the NIDDK Anniversary Scientific Symposium on the NIH campus and the 60th Anniversary Celebratory Dinner later that evening at the Bethesda North Marriott Hotel and Conference Center.

IV. ANNOUNCEMENTS
Dr. Stanfield

Confidentiality

Dr. Stanfield reminded Council members that material furnished for review purposes and discussion during the closed portion of the meeting is considered confidential. The content of discussions taking place during the closed session may be disclosed only by the staff and only under appropriate circumstances. Any communication from investigators to Council members regarding actions on an application must be referred to

the Institute. Any attempts by Council members to handle questions from applicants could create difficult or embarrassing situations for the members, the Institute, and/or the investigators.

Conflict of Interest

Dr. Stanfield underscored that advisors and consultants serving as members of public advisory committees, such as the NIDDK National Advisory Council, may not participate in situations in which any violation of conflict of interest laws and regulations may occur. Responsible NIDDK staff shall assist each Council member to help ensure that the member does not participate in and is not present during review of applications or projects in which, to the member's knowledge, any of the following has a financial interest: the member, or his or her spouse, minor child, partner (including close professional associates), or an organization with which the member is connected.

Dr. Stanfield noted that, at Council meetings at which applications are reviewed in groups without discussion, that is, "en bloc" action, all Council members may be present and may participate. The vote of an individual member in such instances does not apply to applications for which the member might be in conflict. Regarding multi-campus institutions of higher education, Dr. Stanfield pointed out that an employee may participate in any particular matter affecting one campus of a multi-campus institution of higher education, if the employee's financial interest is solely employment in a position at a separate campus of the same multi-campus institution, and the employee has no multi-campus responsibilities.

To ensure that a Council member does not participate in the discussion of, nor vote on, an application in which he/she is in conflict, a written certification is required. A statement is provided for the signature of the member, and this statement becomes a part of the meeting file. Dr. Stanfield directed each Council member to his or her folder containing a statement regarding conflict of interest in his or her review of applications. He asked each Council member to read the statement carefully, sign it, and then return the signed statement to NIDDK prior to leaving the meeting.

V. REPORT FROM THE NIDDK DIRECTOR
Dr. Rodgers

Changes in Reprogramming Procedures

Dr. Rodgers reminded the Council of two changes in the process for reprogramming funds among budget mechanisms--a topic that he had discussed at the February 2010 Council meeting. First, to move funds among budget mechanisms, the NIDDK and other Institutes will need to prepare reprogramming requests with much greater specificity. With the exception of research project grants, these requests must address subsets of budget mechanisms rather than overall totals for the major categories. For example, for a formal reprogramming of $500,000 into or out of the research training mechanism, details would need to be provided at the sub-mechanism level about specific programs to

be affected, for example, institutional or individual National Research Service Awards. Second, based on the current understanding of the Antideficiency Act, the Institute will be restricted from moving funds in either direction between external budget mechanisms (such as extramural grants and contracts) and internal budget mechanisms (such as the Intramural Research Program and the Research Management and Support category) without notifying the Congress 15 days in advance.

Moreover, in April 2010, the Government Accountability Office determined that reprogramming violations can constitute violations of the Antideficiency Act. In those cases, agency heads will be required to report such violations to the President and the Congress. The NIH Office of Budget has advised NIH components to submit to central NIH by June 15, 2010, all requests for reprogramming funds through September 30, 2010, in order to permit adequate time for consideration by the Department, the Office of Management and Budget and the Congress. To meet these new requirements, the NIDDK will carefully monitor the need for budget adjustments.

Appropriations Process

The NIH Director, Dr. Francis Collins, has testified on the President's Fiscal Year 2011 budget request for the NIH before the House and Senate Subcommittees on Labor, HHS, Education and Related Agencies, on April 28 and May 5, respectively. In his testimony, Dr. Collins recounted recent NIH-funded discoveries and advances, the pursuit of unique opportunities made possible with funds provided by the American Recovery and Reinvestment Act (ARRA), and examples of NIH research priorities for Fiscal Year 2011. Regarding Fiscal Year 2011 success rates, Dr. Collins noted at the House hearing that they are projected to be on the order of 15 percent. This projection takes into consideration inflationary increases in research costs, which are expected to exceed the funding increase in the proposed budget.

Dr. Rodgers said that he was pleased to accompany the NIH Director for his House testimony, along with the NIAID Director, Dr. Tony Fauci, and the NIMH Director, Dr. Thomas Insel. Dr. Rodgers was given the opportunity to respond positively to several questions, including one related to comparative effectiveness research. He also elaborated on the success of partnering with non-governmental organizations to translate the important research discovery that increased physical activity can prevent or delay the onset of type 2 diabetes in those at risk. These findings emerged from the NIDDK's Diabetes Prevention Program clinical trial and follow-up study. In a pioneering step for the health insurance industry, the United Health Group has launched a diabetes prevention and control alliance in partnership with the YMCA and Walgreens to translate these scientific findings to community settings by delivering the interventions in a cost-effective way. The United Health Group will cover these services at no charge to participants in its health plan in six cities. This partnership is a milestone in evidence-based diabetes prevention programs.

Dr. Rodgers noted that, after deliberations and possible changes at the subcommittee and full committee levels, House-passed and Senate-passed versions of the appropriations bill

are typically reconciled in a conference bill that is then passed in identical form by both chambers and sent to the President for signature before the end of the Fiscal Year. Dr. Rodgers reminded the Council that the President's Fiscal Year 2011 budget request for the NIDDK is approximately $50 million above the Institute's Fiscal Year 2010 enacted budget level--about a 2.6 percent increase. These figures are exclusive of funding under the American Recovery and Reinvestment Act (ARRA), but inclusive of funds for the wide Special Statutory Funding Program for Type 1 Diabetes Research, which the NIDDK manages.

American Recovery and Reinvestment Act (ARRA)

Dr. Rodgers reported that, to date, the NIDDK has obligated or committed a total of $365.7 million or 82 percent of its $445.4 million in ARRA funding. He noted that the NIDDK plans to spend the remaining funds on a number of activities including R56 awards, R24 grants, RC4 grants, and administrative supplements. He reminded the Council that the two-year ARRA funding program ends September 30, 2010.

Special Statutory Funding Program for Type 1 Diabetes Research

Bills have been introduced in both the House and Senate (HR. 3668 and S. 3058, respectively) to extend the Special Statutory Funding Program for Type 1 Diabetes Research from 2012 through 2016, along with a parallel diabetes prevention and treatment program for Native Americans administered by the Indian Health Service (IHS). Absent this legislation, these programs are slated to end September 30, 2011. The bills would provide each agency with a total of $1 billion over the five-year period covered. There is already considerable co-sponsorship of the bills, which were introduced by Rep. Diana DeGette (CO) and Senator Byron Dorgan (ND).

VI. NEW OPPORTUNITIES FOR BEHAVIORAL AND SOCIAL SCIENCE RESEARCH AT NIH
Dr. Gregory Germino, Deputy Director

Dr. Germino described a trans-NIH initiative--the Basic Behavioral and Social Science Opportunity Network--OppNet (http://oppnet.nih.gov). Basic behavioral and social science research is defined as research that furthers understanding of fundamental mechanisms and patterns of behavioral and social functioning relevant to the Nation's health and well-being, and as they interact with each other, with biology, and the environment.

Goals of OppNet and Its Importance to NIDDK

The goals of OppNet are to support activities and initiatives to focus on basic mechanisms of behavior and social processes and to expand NIH funding for this research area. There are three major categories of research addressed by OppNet: (1) behavioral and social processes--e.g., group processes, learning, social cognition, emotion/motivation, (2) biopsychosocial research, including the study of interactions of

biological factors with behavioral or social variables, such as genetic-environmental interactions with behavior, psychosocial stress and disease; the interaction of circadian sleep rhythms and behavior; and the effects of social networks on the spread of disease, and (3) data collection, modeling, and research design, including the development of better methodologies and tools.

Dr. Germino noted the importance of briefing the Council on this initiative because of its trans-NIH scope, the impact, albeit modest, it will have on NIDDK's planed funding, and the likelihood that the Council will need to review some projects for potential funding. Importantly, NIDDK grantees need to be made aware of the program and encouraged to avail themselves of the research and funding opportunities it presents.

OppNet as a Means of Addressing Poor Health Behaviors

Dr. Germino pointed out that OppNet can be an important means of addressing poor health behaviors. Research has shown that health in the U.S. is influenced by factors in five domains: genetic predisposition, social circumstances, environmental exposure, health care, and behavioral patterns. Behavioral patterns account for approximately 40 percent of the increased risk for premature death (Schroeder, et al. We Can Do Better Improving the Health of the American People. *NEJM* 357:1221-1228, 2007). Therefore, even optimal health care for the entire U.S. population would have only a modest impact on premature death--absent changes in behavior. Research studies have already demonstrated effective behavioral approaches to reduce the risk of type 2 diabetes, obesity and hypertension--key research areas supported within the NIDDK mission. However, the American public and its health practitioners are not yet fully embracing sound, science-based advice for healthy behaviors, such as moderating food intake, increasing physical activity, and reducing salt in the diet.

Dr. Germino noted that the gulf between the availability of evidence-based proof-of-efficacy for behavioral interventions and their successful clinical implementation in the real world can be described as a "second valley of death." The first "valley of death" is the gulf that exists between demonstrating "proof-of-concept" for an intervention in the laboratory and then bringing it to a clinical setting where it can directly benefit patients. Both of these gulfs are "valleys of death" in translational research and medicine.

The NIH has long supported different approaches to bridge the second translational gulf on the research continuum, including educational programs for patients and medical/health practitioners. The NIDDK has contributed to these efforts through such programs as the National Diabetes Education Program, the National Kidney Disease Education Program, and the Institute's three national information clearinghouses. In the same vein, the NIH has supported behavioral research, including the study of approaches that can help motivate people to follow treatment guidelines. Yet, for behavioral approaches, there has not yet been optimal implementation of evidence-based proof-of-efficacy into clinical practice.

The OppNet program is aimed at addressing this translational issue by broadening the scope of behavioral research to include a new focus on studies that will provide a deep understanding of behavioral and social processes. Just as an increased understanding of the fundamentals of biology can inform disease-focused research, so too can insights into fundamental behavioral and social processes inform the adoption of behaviors that have been demonstrated by research to be effective in improving human health.

Funding of OppNet

Dr. Germino described the funding expectations for the trans-NIH OppNet program and the impact that NIDDK's participation will have on the Institute's budget. The estimated total funding for this program is $12 million in FY 2010, $20 million in FY 2011, and $30 million for each of the Fiscal Years 2012 through 2014. In FY 2010, $10 million will come from Office of the NIH Director from ARRA funds and $2 million from HIV-AIDS specific funds. From Fiscal Years 2011 through 2014, funding will be from the various NIH components that support extramural research. Contributions will be calculated as a percentage of each organization's base appropriation. The NIDDK will contribute about $1.2 million in FY 2011 and $1.8 million annually from FY 2012 through FY 2014.

For FY 2010, the NIH will fund meritorious applications that have been received in response to a series of already published funding opportunity announcements, including competitive revisions to existing grants and short-term training opportunities that were consistent with the time-sensitive use of ARRA funds. For FY 2011 and beyond, the NIH has solicited ideas for new initiatives from both within and outside of the NIH community. For example 320 responses were received in response to a Request for Information for web-based input. Based on the response, the NIH is now developing funding opportunity announcements for FY 2011 that will be posted by mid-summer 2010, with a late fall submission date. In addition, the NIH will sponsor a large conference in the fall of 2010 to obtain broad input for the development of OppNet plans for FY 2012 and beyond.

In closing, Dr. Germino encouraged the Council members to review the OppNet website and to inform their colleagues about the program. He acknowledged the efforts of NIDDK staff members Drs. Christine Hunter, Sue Yanovski and Phil Smith in representing the NIDDK on the various committees guiding OppNet's development.

Questions and Discussion

Will OppNet be funded with new, additional funds provided specifically for the program, or must program costs be absorbed within existing budgets? Dr. Germino replied that funding will be from existing budgets. That is one reason that NIDDK grantees need to be aware of the program and take advantage of its funding opportunities.

What will NIH do to overcome human resource issues with respect to obtaining the broadest possible range of research applications and having the staff expertise to address

them? What steps will NIH take to reach out to the full community of behavioral research and social science researchers--many of whom may not be engaged in health research or familiar with NIH funding announcements and processes? Dr. Germino replied that there has been extensive discussion regarding these issues among NIH staff members, the behavioral and social science communities, and their advocacy groups. The NIH is seeking very broad representation and participation in this initiative. For example, one goal for the fall meeting is to engage a wide spectrum of individuals from many relevant disciplines, including behavioral economics, so that the program is developed in a comprehensive manner.

VII. ADVISORY COUNCIL FORUM: Part 1

A. "NIDDK Informatics Concept Development"

Dr. Ronald Margolis, Senior Advisor for Molecular Endocrinology, Division of Diabetes, Endocrinology and Metabolic Diseases

Dr. Margolis' presentation focused on the NIDDK's efforts to integrate data in ways that might foster an evolution to a more comprehensive informatics grid. The NIDDK is striving to facilitate data sharing, especially among some of its basic science consortia. The Institute is also exploring other data-sharing and informatics initiatives underway at NIH and the Department of Health and Human Services in order to align itself better with these activities in a broader context.

Dr. Margolis provided an overview of efforts to evaluate and develop informatics at NIDDK, including the formation of a small Informatics Working Group of Program Directors within the Division of Diabetes, Endocrinology and Metabolic Diseases (DEM). The Working Group is trying to address several issues related to the need for informatics support of the Institute's mission. These issues include the integration and improved usability of throughput data accrued through NIDDK research; the linking of basic science discovery projects to enhance their impact; the feasibility/practicality of providing broader availability of some clinical data derived through NIDDK research; and the kinds of bioinformatics tools that might be necessary to enhance or facilitate hypothesis-generation and testing. To guide its deliberations, the Working Group has focused on a few key questions: What types of data are useful and how should they be presented? Who are the NIDDK stakeholders with respect to contributing to and using these data? How can the NIDDK assess and enhance the utility of these data? How can the access to and mining of these data be maximized with the ultimate goal of fostering greater research translation?

The Working Group began by looking at several of NIDDK's existing basic science consortia: the Beta Cell Biology Consortium (BCBC; www.betacell.org), the Nuclear Receptor Signaling Atlas (www.NURSA.org), the Mouse Metabolic Phenotyping Centers (www.mmpc.org), and the Animal Models of Diabetic Complications (www.amdcc.org). These programs develop large amounts of genomics, proteomics, metabolomics, and phenotypic data--mostly in the mouse. The questions are: How can NIDDK tie the data in these programs together so that it is easy for investigators to go from one data source

to another to facilitate their own research? From these programs, can NIDDK create a data "mart" as an adaptable, extensible network that will help foster evolution to a more comprehensive informatics grid useful to stakeholders? If NIDDK can accomplish these first steps, a further step might be the development of a community-based network or grid to foster data integration across the NIDDK universe of diseases, investigators, and other potential users. Such a network could serve not only as a data "mart," but also as an access point for resources/tools and a catalyst for discovery.

To pursue these ideas, the Institute has formed an NIDDK Consortium Interconnectivity Network--dkCOIN. Work is now underway to develop a proof-of-concept for the integration and widespread utility of data across the four basic science consortia cited. The Institute is planning a workshop in September 2010 to evaluate lessons learned from the dkCOIN efforts, and to consider extending them to a larger network or grid. To that end, the NIDDK is consulting with other informatics efforts across the NIH and in the public and private sectors. Examples include NCI's caBIG (https://cabig.nci.nih.gov), NHLBI's Cardiovascular Research Grid (http://cvgrid.org/), the Clinical Translation and Science Award Consortium (http://www.ctsaweb.org/), and the Biomedical Informatics Research Network (http://www.birncommunitiy.org).

Dr. Margolis concluded by thanking the other members of the DEM Informatics Working Group: Kristin Abraham, Olivier Blondel, and Art Castle. The Working Group is now being expanded to include representatives from the other extramural divisions. He also expressed appreciation for input from the following individuals associated with NIDDK consortia: Drs. David Steffen and Neil McKenn (NURSA); Dr. J-P Cartailler (BCBC) and Dr. Rick McIndoe (MMPC/AMDCC).

B. "Informatics: The National Center for Biotechnology Information (NCBI) and the Genetic Testing Registry"
Dr. James Ostell, Chief, Information Engineering Branch, NCBI, National Library of Medicine, NIH

Dr. Ostell earned a Ph.D. in molecular biology from Harvard University, then developed commercial software for biotechnology. He came to NCBI when it was created in 1988. As Chief of the Information Engineering Branch, Dr. Ostell has been responsible for designing, developing, building, and deploying all the production resources at NCBI from its beginning--both through his own efforts and by directing and coordinating the efforts of a large and rapidly growing group of talented NCBI staff. These resources include PubMed, GenBank, BLAST, Entrez, RefSeq, dbSNP, PubMed Central, dbGaP, and many others. Under Dr. Ostell's direction, the Branch has grown to become a flagship public resource in bioinformatics. In 2007, Dr. Ostell was inducted into the National Academies, Institute of Medicine.

Dr. Ostell pointed out that NCBI faces the same issues that NIDDK is considering today: How can an organization be effective in its bioinformatics efforts and how can it measure whether it has accomplished its goals? He said that informatics is neither trivial nor magical, although it is often perceived as one or the other. Instead, it is like laboratory

science in that it is partly vision, but it also requires doing lots of hard, detailed work; finding and recruiting talented people; asking the right questions at the right time; and knowing when the answers emerge.

Background on NCBI

The NCBI was established in 1988 by Public Law 100-607. It was intended to create automated systems for knowledge about molecular biology, biochemistry, and genetics; to perform research into advanced methods of analyzing and interpreting molecular biology data; and to enable biotechnology researchers and medical care personnel to use the systems and methods developed. Over time, NCBI has been the builder and provider for several databases including GenBank, Blast, PubMed, dbSNP, dbGaP, and RefSeq. It also supports other resources such as OMIM, GeneTests, and Gene Reviews, and provides basic research and training in computational biology.

The NCBI currently has about two million users a day on its website and about 5,000 web hits a second, which makes its site the most heavily used in bioinformatics. Within the federal government, the use of the NCBI website is second only to that of the Internal Revenue Service website at tax time. The NCBI delivers about 8 terabytes of data per day to external users. The number of users has risen over time--beyond the envisioned user community of molecular biologists. That continuing growth in users probably reflects a greater use of molecular resources across disease disciplines; the rise of informatics science and the use of high throughput technology; an increase in the number of people who consider themselves to be doing biomedical research; and the interest of the general public.

Establishment of Genetic Testing Registry

To illuminate some of the challenges of bioinformatics, Dr. Ostell recounted the establishment of the Genetic Testing Registry (GTR). In announcing GTR in March 2010, the NIH said that: "… there is no single public resource that provides information about the validity and usefulness of [genetic] tests. NIH believes that easy access to such information is vital to facilitate research and to enable informed decision making by patients, caregivers, health care providers, payers, and policy makers. Therefore, NIH is initiating the development of the GTR, an online resource that will provide a centralized location for researchers, test developers, and manufacturers to voluntarily submit information about genetic tests such as their intended use, validity, and utility."

Dr. Ostell underscored the difficulties of obtaining voluntary participation, and the tremendous importance of focusing on the use, validity, and utility of genetic information. He then laid out the basic fact-finding paradigm for the case of a variation in the genome that can be measured.

- *Analytical Validity:* How can the genetic variation be accurately assayed? What is the variation and where is it located?

- *Clinical Validity:* Does the assay accurately predict/diagnose disease? What is the effect of the genetic variation, and is there convincing evidence that the effect is real? Is there more than one phenotype involved? Is there merely an association between the genetic variation and the phenotype, or is the variation causative?

- *Clinical Utility:* How can this genetic knowledge benefit the patient? Is there an intervention that can help a patient who is told he has a high risk for a particular genetic disease? Also, can this information contribute to clinical effectiveness research?

Landmark, Coodinate and Address-Based Systems

Dr. Ostell described how analytical validity (the basic facts as to the "where and what" of the genetic variation) can be pursued using a landmark-based data system (Online Mendelian Inheritance in Man – OMIM), a coordinate-based data system (dbSNP), or an address-based data system (RefSeqGene).

An example of a landmark-based data system is the Online Mendelian Inheritance in Man (OMIM) system--an online resource for classical findings in medical genetics. It also contains some reports of mutations that have been associated with disease. For example, this database provides information about the deletion of certain base pairs in exon 10 that leads to deletion of phenylalanine, codon 508. Dr. Ostell likened this to a local neighborhood resident giving directions to a visitor by saying: "Go to the big tree and go over the hill." People living in the neighborhood--just like scientists working on a particular gene--know what this type of landmark description means, but for others who are not in the local community, it is not particularly helpful. Moreover, local landmarks easily change; trees get knocked down and new genetic information can change the location of exon 10. For those living in a local community or working in discrete area of genetics, these types of landmark designations may retain meaning even when changed; however, for outsiders, they can be confusing or meaningless. Landmark-based data are typical of clinical genetics because much of these data was acquired in gene-by-gene Mendelian fashion, before the advent of high throughput thinking.

Limitations also exist in a coordinate-based system such as dbSNP. Using this database is somewhat similar to using a global positioning system (GPS) on a car. The system will provide location information such as: N 37 degrees 43.69 minutes, W 97 degrees 28.39 minutes. However, there is no context as to whether the location is on land or water, or in the mountains, the city or the country. However, as Dr. Ostell pointed out, the analogy to a GPS is imperfect. Because the universe of known genetic information is continuing to increase in size, and every human being is different, the human genome map continues to have shifts. Importantly, small shifts in genetic coordinates can cause mathematical problems when computers are used to mine vast amounts of data in very large informatics systems.

Dr. Ostell noted that, as a partial solution to these problems, the NCBI introduced the RefSeqGene system. RefSeqGene, a subset of the Reference Sequence (RefSeq) project,

defines genomic sequences of well-characterized genes to be used as reference standards. It is a genomics coordinate system that focuses on chunks of genomic DNA that surround a gene. These gene-chunks are independent of the major genetic annotations on the large genome map, which is subject to change. In other words, the chunks of genetic sequence are not numbered from exons. Metaphorically, they are can be considered "islands" of genetic sequence that "float" on a changing public map of the human genome. Yet, their internal coordinates provide a stable, medical genetics system. When the NCBI creates a record in RefSeqGene, it goes back to the research community that works on the gene to ask investigators about the sequence they are using. The NCBI then tries to make that implicit "landmark" system explicit in RefSeqGene. The result is more of an address-based system in which the location provided is something like: 253 Center Street, Wichita, Kansas, USA. There is a coordinate, but also a context. The information is meaningful, not only to the local research community, but also to a very broad audience. The NCBI developed the RefSeqGene system in conjunction with the College of American Pathologists, and it is now being expanded into an international system.

Dr. Ostell noted that the NCBI does what it calls "information engineering," a process that uses mathematics, methods of aligning sequences, and other techniques to bridge gaps as the field of genetics moves into the genomic era. For example, the NCBI uses computational techniques to align whole sequences with the public human genome map, and thereby connect that information to a research coordinate system. The NCBI is linking information on allelic variants from the OMIM database to information on single nucleotide polymorphisms in the dbSNP database, which is considered the comprehensive catalogue of genetic variation. The Center is striving to expand medically important landmarks in dbSNP. The NCBI supports GeneReviews, which are curated reviews of genetic diseases written by experts, and it also provides computer support and an electronic authoring system. When a structured electronic document comes to the NCBI, the data it contains can be exported to update information in an existing database. The NCBI is also part of the genome editorial board, which makes necessary corrections in the public genome map in conjunction with other groups.

New Emphasis on Clinical Data

The NCBI is importing more data from clinical databases, such as those maintained by health care systems. These systems can have important information about links between genetic mutations and diseases, and other insights derived from clinical testing performed in a clinical practice setting. For example, by working with clinical databases, the NCBI is collecting data on how often certain genetic deletions have been seen and how often they have been called pathogenic. The NCBI centralizes this information so that its users can see patterns of recurring deletions and mutations. Also, when the NCBI knows the genetic probes being used in a database platform, it can place them on the human genome map via an alignment process. If it finds mistakes, the NCBI can encourage the platform maker to make necessary corrections. The NCBI is launching a new database, ClinVar, which will take clinical information from various sources and aggregate the data into groups of clinical significance at the levels of the gene, the protein, the transcript, the genome, and, ultimately, some clinical condition or phenotype. ClinVar will connect

genotypes to phenotypes, linking data from the research and clinical practice sides in an aggregated manner. One benefit will be that labs registering a genetic test will not have to explain a mutation repeatedly; rather, they will be able to point to a record in ClinVar to demonstrate clinical validity. This database will thus reduce the reporting burden on participants.

NCBI Perspectives

The NCBI makes it a point to ensure that all its data are publicly available and also encourages the developers of other informatics databases to do the same. For example, the NCBI works to garner information from locus-specific databases, most of which are privately maintained. The Center has set up a mechanism to import such data, convert it to standard nomenclature, help in its validation, map mutations onto RefSeqGene, link the data to relevant publications, and provide attribution for the data to the originating locus-specific database. The NCBI also hosts locus-specific database tools that individuals and groups, such as disease-focused patient advocacy groups, can use free of charge to build their own databases, provided that they agree to make the data publicly available. In addition, the NCBI may contact NIH institutes or other organizations that operate databases to see if their data can be made more widely accessible via the Center. Dr. Ostell mentioned specific efforts to work with the NIH Office of Rare Diseases and the National Institute of General Medical Sciences. This partnering approach--a "coalition of the willing"-- also helps the NCBI expand the scope of its operations, because the Center recognizes that it cannot possibly represent every specialty or specialized resource that exists. Therefore, a record in dbSNP or other NCBI data system may point the user back to an originating source that provides detailed pharmacokinetics or other data not maintained by the NCBI.

In addition to trying to pair the large, centralized resources of the NCBI with specialized resources, the Center also weighs duration issues in considering its undertakings. The NCBI chooses to initiative activities that are expected to continue over decades, such as PubMed, because the Center's operations receive relatively stable funding directly through the NIH budget, rather than through grant awards that have short timeframes and must be renewed. However, the NCBI tries to work with grant-funded informatics efforts to encourage standardized identifiers, and the use of NCBI's centralized resources for the importation or linking of such data. When the NCBI makes the decision to start a new activity, it does not consider the undertaking as an isolated project. Rather, the Center leadership thinks about how the new activity will fit into the existing set of NCBI undertakings. The NCBI seeks to leverage resources by building upon and integrating existing efforts.

Questions and Discussion

A Council member commented that there are several factors contributing to the NCBI's success: (1) the NCBI has stable funding, which enables it to undertake long-term activities, (2) the NCBI has a strong partnership with the European Bioinformatics Institute, which gives its efforts a worldwide scope, and (3) the NCBI is not disease or

Institute specific. The NCBI's efforts are meritorious and laudable. A goal for NIH should be to create such central repositories that are available to everyone, and then allow specific Institutes or research communities to build a "front end" to customize the data for their specific uses. This would be preferable to each Institute having its own informatics strategy and operations. However, the reality is that NCBI is underfunded relative to the Institutes.

The Genetics Testing Registry seems to be a departure from previous types of NCBI activities. The technology is relatively straightforward and, in some ways, simpler than working at the genomic level. However, the proposed NCBI linear model--in terms of clinical validity and clinical utility--does not appear to speak to the complexity of interpreting genetic testing. There will be a need to provide oversight in terms of the data that are being deposited in the registry because there is an inordinate amount of subjectivity and a wide variation in sensitivity and specificity. Misuse or misinterpretation of the data could occur. Dr. Ostell commented that the Genetic Testing Registry was not initiated by the NCBI, but rather, by the Office of the NIH Director. The NCBI recognizes and is willing to accept the challenges this undertaking presents, and believes that, by addressing them in a transparent manner, the registry provides an opportunity to promote improvements in the field. The NCBI also sees a responsibility for the NIH to build upon the sequencing of the human genome, which represented a substantial investment of NIH funds. The first stages of the Genetic Testing Registry will focus on relatively straightforward genetic diseases such as sickle cell disease, for which the underlying genetic mutation is known. Next will be diseases involving multiple mutations, such as cystic fibrosis. Then, the registry will move on to diseases involving alleles. Beyond that are health conditions that reflect multigenic effects. Dr. Ostell said that it is possible for the Genetic Testing Registry to identify existing diagnostic tests, some of which are already available commercially, without stating that the tests have clinical validity. The test developer can provide published articles and other support for the test that the data user can consult. Dr. Ostell indicated that it is NCBI's intention that every genetic test in the registry will have an accession number that can referenced in the published literature by researchers whose findings either support or refute the test. This transparent approach and iterative process will be the best driver of accuracy. The entry of some major tests into the registry will encourage other test makers to join the effort. Dr. Ostell believes that the NCBI needs to take the first step, even though it doesn't know all the answers.

There is an onslaught of new information being produced by high throughput methodologies not only in genetics, but also in other fields. In this context, how should advisory groups such as the Council proceed? Dr. Osteen recommended that members of advisory groups be skeptical about proposed initiatives and consider them as carefully as they would a new activity to be undertaken in their own laboratories. The NCBI tends to start with straightforward examples and then take the next "obvious" step because it is close enough to the real world to provide a sense of utility. In other words, if the step is taken, there is a likely result that will help researchers. Moreover, resourceful people will tend to find even more uses than anticipated. Focusing on what can be realistically

accomplished is often better than being too forward-looking or ambitious. It is important to recognize that it is not possible to do everything.

What does the NCBI know about its two million website users? Can the NCBI capture their navigation of the website and how they are using the data? Can the NCBI obtain feedback from its users? Dr. Ostell noted that, because people share information, there are probably more users of the NCBI's website information than the two million documented annual users. Initially, the NCBI did not attempt to analyze user patterns, but more recently, it is ramping up such efforts because they can lead to improvements in NCBI operations that will benefit users. The NCBI has developed a login system--MyNCBI--through which a user can collect information and store preferences. The Center does not look at what individuals are doing. Rather, it looks at overall statistics, including which of the databases are used most heavily. Just like Google and Amazon, the NCBI recognizes that the way a concept is worded can have a huge impact on whether people click on a pulldown or link. The NCBI believes that its investment in a large database to analyze the log is worthwhile because the resulting small improvements the NCBI can make will affect thousands of users. He cited an example in which the Center recognized certain gene-search patterns among website users that then enabled the NCBI to help direct those types of users to a database that was highly relevant to their interests.

How is the NCBI prioritizing activities--both ongoing and planned--in this time of limited resources to ensure that it does not lose momentum? Does the NCBI need a more proactive strategy to determine what undertakings are needed? Will some opportunities to combat disease be lost because the NCBI is not moving fast enough? Dr. Ostell replied that, structurally, the NCBI developed from a tiny group of NIH staff members who were working within the National Library of Medicine. As the NCBI has grown, it has had a succession of planning processes to build on that initial effort. Dr. Ostell said that, in NIH terms, the NCBI is the laboratory he directs as a research scientist. Therefore, like other NIH intramural activities, the NCBI has a Board of Scientific Counselors, which is an external advisory group that meets every six months to review the Center's activities, priorities and initiatives. In that regard, this Board acts in a way that is functionally similar to a National Advisory Council. Recognizing that the NCBI serves a very broad community, the NIH has also established an NCBI Resource Board to guide the Center's activities. Composed of several Institute Directors and the NIH Director, this Board reviews priorities and funding. The NCBI also benefits from informal input, such as that received from presentations at National Advisory Council meetings and in other venues. The Center is also very receptive to speaking with Institutes and Centers about new ideas. That is the way that the NCBI became an active participant in the NIH Genome Wide Association Studies (GWAS) initiative, which was initially spearheaded by the NHLBI and the NHGRI. The NCBI also receives high levels of public feedback. Dr. Ostell said that the NCBI welcomes all types of feedback because comments from users and other parties reflect interest in the Center's activities and can lead to operational improvements.

VIII. ADVISORY COUNCIL FORUM: Part 2 - "Data Sharing Initiatives within and around the Clinical and Translation Science Award (CTSA) Consortium"

Dr. Michael Kahn, Associate Professor of Pediatrics and Co-Director of the Clinical and Translational Sciences Institute at the University of Colorado, Denver; Director of the Division of Clinical Informatics at the Children's Hospital, Aurora, Colorado; and Co-chair of the Clinical and Translational Award (CTSA) Consortium's Key Functional Committee on Bioinformatics

Dr. Kahn's research interests include real-time clinical decision support linked to clinical outcomes monitoring; clinical data warehouses for both operational and retrospective research support; integration of electronic medical records with prospective research; and translational research informatics for both "bench to bedside" and "bedside to community" translational settings. Prior to his current positions, Dr. Kahn was a member of the faculty in the Departments of Medicine, Computer Sciences, and Biomedical Engineering at Washington University School of Medicine, St Louis; the Director of Advanced Clinical Systems at BJC Health Systems; and a member of the commercial clinical trials software industry. Dr. Kahn received a B.S. in Biological Sciences and a B.A. in Chemistry from the University of California, Irvine, and an M.D. from the University of California, San Diego. He received a Ph.D. in Medical Information Sciences from the University of California, San Francisco.

Dr. Kahn noted that his presentation would provide very general background information regarding data sharing and informatics. His objective was to point the audience to examples of resources from which more detailed information can be obtained. He thanked individuals who provided material, and, throughout this presentation, he pointed out the organizations that are leading and/or participating in specific data sharing efforts.

Integration of Different Perspectives and Types of Data

For context, Dr. Kahn described the concept of the "circle of integration," put forward in a schematic from the Ohio State Medical Center. This circle depicts the effort to integrate data so that it can be better managed, analyzed and used. The circle represents efforts to bring together in new, creative and synergistic ways the knowledge that is being gained from many sources, including phenotypes in patient populations, findings in the scientific literature, results of biomarker research, and functional/mechanistic work from basic science labs. This is a new era of integrated translational research informatics.

Dr. Kahn noted that communications among biomedical scientists and informatics experts can sometimes be difficult because they have different perspectives. For a biomedical scientist, his or her hypothesis-driven research is the "science," while informatics is the "plumbing." However, for a person working in the informatics field, the science is in the informatics itself, while biomedical research findings provide interesting examples of data that informatics can illuminate. It takes both perspectives to optimize data integration and management.

Centralized vs. Distributed Data Sharing Systems

Dr. Kahn said that, when discussing data sharing, it is important to consider the use of central vs. distributed data sharing models. The centralized approach is the traditional model, which is used by the National Center for Biomedical Information (NCBI). Moreover, many clinical trials use this traditional model in which participating clinical research centers periodically submit their study data to a central data coordinating center, which is the only place that maintains and has full access to all the data.

In the clinical world, the centralized model is now being replaced by a distributed data sharing model. In the distributed model, a local site that owns data continues to own and store it locally, but decides to make portions of the data available to a central site on a need-to-know or as-needed basis--typically in response to authorized data queries. The locus of control for access and administration rights to the data rests with the local site. There is a retention of local control coupled with the granting of shared access. The difference in the centralized and distributed models is a transformation in geometry and control. When people describe distributed models or networks as "grids" or examples of "grid computing," they generally mean that these models function like the Internet by providing an infrastructure or architecture that enables large-scale, distributed data sharing, with important features such as interoperability and data security.

In general, data sharing systems operate in well-defined communities, which also make efforts to facilitate data sharing with the public. There is usually an established infrastructure and techniques for data sharing, and a means to mediate data integration across data sources. In both centralized and distributed models, data sharing can occur either immediately (real-time mode) or intermittently (batch mode). In distributed data sharing systems, security and confidentiality issues are critically important, such as how to identify/register the user and his institution, and how to determine the activities (such as authorized queries) that a user will be permitted to initiate.

Examples of Distributed Data Sharing Systems and Approaches

Dr. Kahn then presented several examples of distributed data sharing oriented to biomedical research.

- *Biomedical Informatics Research Network – BIRN:* One of the oldest networks, BIRN is supported by the NIH's National Center for Research Resources at the University of Southern California. BIRN has tried to create a generalizable infrastructure, which can then be leveraged by specific research communities to meet their own data needs. BIRN provides a user-driven, software-based framework for research teams to share significant quantities of data--rapidly, securely and privately--across geographic distance and/or incompatible computing systems. Groups may choose whether to share data internally or with external audiences. In either scenario, hardware and data remain under the control of individual user groups. Dr. Kahn

considers BIRN to be oriented primarily toward the basic sciences, with a special focus on bioimaging and genetics.
www.birncommunity.org

- *Cancer Biomedical Informatics Grid – caBIG:* Dr. Kahn described caBIG as a centralized, top-down network, which is supported by the National Cancer Institute. Composed of 1,500 participants (researchers, physicians, and patients) from 450 organizations, caBIG is a collaborative information network to accelerate the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer--with the goal of ultimately improving patient outcomes. The network connects scientists and practitioners through a shareable and interoperable infrastructure; develops standard rules and a common language; and builds or adapts informatics and analytic tools. The scope of caBIG is now evolving to encompass other diseases in addition to cancer. In related efforts, the Ohio State Medical School's Center for IT Innovations in Health Care and its CTSA program are developing a *Translational Research Informatics and Data Management Grid (TRIAD)* with NIH ARRA support. TRIAD is extending cancer data-sharing tools to the much broader translational research activities of the CTSAs. Both caBIG and TRIAD are also moving beyond data sharing to resource sharing, such as the sharing of analytic tools.
www.cabig.nci.nih.gov

- *Informatics for Integrating Biology and the Bedside - i2b2:* Harvard's i2b2 Center is one of several National Centers for Biomedical Computing funded under the NIH Roadmap initiative. Building heavily on work done by Harvard's Partners in HealthCare system, the i2b2 Center is developing a scalable informatics framework that enables clinical researchers to use existing clinical data for discovery research. When combined with genomics data, the informatics framework will help to facilitate the design of targeted therapies for individuals with diseases having genetic origins. Harvard's i2b2 Center has worked closely with the main hospitals within the Harvard "Catalyst" CTSA (Massachusetts General, Brigham and Women's Hospital, Beth Israel Deaconess Medical Center, and Children's Hospital, Boston) to facilitate their adoption of i2b2 instances as a vehicle for conducting clinical research at these heterogeneous biomedical research institutions, and as a prelude to sharing aggregate data by means of a web-based, federated query system known as Harvard's Shared Health Research Informatics Network – SHRINE. The i2b2 software platform is also being used by many other CTSAs and academic health centers, as well as industry. The platform takes internal tools and makes them available in a free, open-source environment. The i2b2 software, called "the hive," consists of independent modules that share a common messaging protocol. The concept is to provide a framework that permits different types of activities to occur through use of these tool sets. For example, there is a repurposing of data in electronic medical records to enable cohort identification and population identification that is useful in designing clinical trials. This approach also enables the dovetailing of phenotype and genotype information. The i2b2 software platform encourages the development of a community of tool

builders and tool improvers because institutions can share their new tools with other users in an iterative process.
www.i2b2.org/

Dr. Kahn noted several examples of the use of the i2b2 software platform:

Harvard Affiliated Hospitals i2b2 Inflammatory Disease Registry : The i2b2 software tools are being used to establish a registry of 3,000-5,000 adult patients and over 300 pediatric patients with inflammatory bowel disease in the Boston area. Institutions participating in this network will share data to answer driving biological questions--an objective that is characteristic of these types of research networks.

Cross-Institutional Clinical Translation Research Project – CICTR: This NCRR-supported translational informatics demonstration project involves three CTSA partners (University of Washington; University of California, San Francisco; and University of California, Davis), along with several collaborators. It is considered by many to be the West Coast equivalent of Harvard's SHRINE. Led by the University of Washington's Institute of Translational Health Sciences, this project is using i2b2 software to explore cohort selection for disease-focused studies across the three CTSAs, with a focus on diabetes and cardiovascular disease. The project involves over 3 million anonymized patient records.
http://www.bhi.washington.edu/by-name/byname.html#6

Childhood Arthritis and Rheumatology Research Alliance Network – CARRANet: This network reflects the cooperative efforts of the Harvard Center for Biomedical Information, the Harvard Medical School and the Children's Hospital Informatics Program, Children's Hospital, Boston. It involves 60 different sites across the country, representing approximately 20,000 pediatric rheumatology patients. Harvard has agreed to host the i2b2 technology for these very small sites because they don't have the technical expertise to operate local i2b2 data sites (nodes). The data will be in a distributed system, but the management of that system will be centralized by Harvard.
www.carragroup.org/

Dr. Kahn pointed out that two of the more clinical-sciences-oriented networks are:

- *HMO Research Network – HMORN:* The HMORN is a practice-based, population-based consortium of 15 very large and geographically dispersed community-based health systems. The network has data on 11 million U.S. citizens. The goal of HMORN is to transform health care practice through population-based research in order to improve the health and health care of broad populations. The HMORN was initiated with ten sites in 1994. The early years were focused on addressing organizational, contractual and other administrative issues. The HMORN has organized itself into interest groups or collaborative groups based on particular diseases, such as cancer and cardiovascular disease. The network also includes data

from outcomes research, and is seeking to add more areas. The FDA has started to look to the HMORN for surveillance work with respect to vaccine safety links. The HMORN model is slightly different from other networks in that none of the data ever leaves the local participating sites. There is a standard model for data extraction, entry and storage at the local sites. The central coordinating system makes some computer programming assistance available to the participating sites.
www.hmoresearchnetwork.org

- *Distributed Ambulatory Research in Therapeutics Network – DARTNet:* Dr. Kahn noted that DARTNet grew out of the Department of Family Medicine at his own institution, the University of Colorado. It is funded by the Agency for Healthcare Research and Quality (AHRQ) and the American Academy of Family Physicians. The network provides to participating sites the ability to extract data and maintain it locally, along with the ability to share it with others. The Network focuses on removing translational barriers, for example, through timely translation of research discoveries to very small medical practices. Data are included on approximately 4.5 million patients seen at 345 practices. Importantly, the network has found ways to incentivize small practices to participate by offering them outcome indicators that can help them improve the quality of their clinical decisions.
www.dartnet.info/

Dr. Kahn underscored that long-term efforts are needed for successful data sharing networks. He estimates that a ten-year commitment is probably necessary for these types of programs. However, starting a network is somewhat easier now than in the past because there are model documents and examples to follow.

Questions and Discussion

How does the "cloud" fit into the development of data sharing networks? Dr. Kahn replied that the term "cloud" is usually used to refer to the Internet. "Cloud computing" is the concept of infinitely scalable access to computing resources or storage. The term essentially refers to a massively elastic, completely endless amount of storage. Dr. Kahn noted that he uses a service on which he can borrow, for a short period of time, an infinite number of computing cycles. For example, if he needs to handle an enormous algorithm, he can have access to 200 processors for two hours for his work and will be charged only for the processing hours. He does not have to invest in ownership of computer equipment to have that capability. Dr. Kahn said that he personally favors the advantages of "cloud computing" and thinks that it has an important and growing role in the evolving computer landscape.

How can data sharing be marketed, particularly to small clinical practices? What kinds of information and analyses are being made available to such practices? Dr. Kahn responded with the DARTNet example. He noted that the American Academy of Family Physicians markets the network opportunity to its members at Academy-sponsored events. The American Academy of Pediatrics is also encouraging its membership to join the network. The types of analyses DARTNet provides to incentivize practices to join

include quality measures for reporting purposes, so that the practices can be in compliance with guidelines regarding diabetes care, influenza vaccination rates, pneumococcal vaccination rates, and other preventive measures. Reports are prepared for the practices on a nightly basis regarding patients coming in the next day, and on a monthly or quarterly basis so that the practices can gauge their performance.

What economic incentives are needed or could be changed to ensure that the great tools of informatics are enthusiastically embraced? Dr. Kahn pointed out that there is already a tremendous incentive for research centers and medical practices to participate in networks because the scope of accessible information permits powerful analyses that might otherwise be infeasible. However, the barriers that exist include concerns about privacy, confidentiality and the discrimination that could occur, especially by the insurance industry, based on disclosure of medical information. Dr. Kahn noted that these barriers are really not an economic issue, but rather, a social issue that needs to be resolved through fundamental policy changes. The Congress is pursuing legislation to address these barriers.

To what extent are the CTSAs involved in funding the databases described? How are the CTSAs coming together as a community to decide which one or ones of these systems to use? Dr. Kahn responded that BIRN was funded before the CTSAs were conceived. The HMORN and DARTNet are primarily AHRQ-funded initiatives. Dr. Kahn said that the major information sharing activity right now is found with i2b2 and TRIAD. However, he emphasized that the CTSA Consortium would not want to be in the position of blessing winners or losers. Rather, the CTSA community is working to ensure that the tools and techniques that are used by any of these networks can be successful. To that end, it is important to have a common vocabulary and standards for data sharing that are independent of any particular networking tool or system. The CTSAs are also spending a great deal of time on regulatory, multi-institutional issues, which go beyond informatics, so that operational agreements can be put in place more rapidly.

How would the informatics approach that NIDDK is taking fit into the existing resources described? What is the wisdom of NIDDK's pursuit of the informatics plans it has outlined? Dr. Kahn responded that there are many factors involved in decision making with respect to informatics. Certainly, there need to be communities of interested parties to do the work, and the NIH Institutes have communities of committed investigators and patient advocacy groups who are eager to capitalize on data sharing capabilities in order to propel research and combat specific diseases. However, it is also important to consider both short-term and long-term needs and opportunities. Immediate decisions should be weighed carefully so that they don't preclude the ability to take advantage of emerging future opportunities. It is crucial to remain aware of what is happening in the overall field of bioinformatics so that organization-specific activities are planned and developed in a way that will enable them to be linked to larger informatics activities, which are continually evolving. In this type of balancing act, incrementalism combined with long-term strategies may be appropriate, so that nimble shifts in direction remain possible as scientific fields and informatics technologies change.

In domains that are oriented toward the clinical sciences, data sharing partners in large institutional settings often will only provide de-identified patient data that do not permit certain research questions to be answered, or that require the questions to be reformulated for application to aggregated data. Has there been any success in resolving this issue? Dr. Kahn said that this issue is still an area of active informatics research. It is recognized that de-identified data is not optimally useful. There is a new informatics subfield looking at anomymization techniques and different ways of re-identifying data so that patient confidentiality and optimal data utility can co-exist. Efforts are ongoing to understand and achieve the appropriate balance--including ways to measure risks, which vary depending upon the size of the population and the query.

Could the issue of de-identification and re-identification of patient data be overcome with resources dedicated to a hardware solution? Dr. Kahn said that he did not believe the solution would be found in hardware that is simply providing encryption functions. The risks of re-identification of data are largely a function of the questions being asked. Some queries may be completely safe in terms of patient confidentiality, whereas others may be compromising. For any given data network, the issue is not a function of the data, but rather, a function of the query.

IX. SCIENTIFIC PRESENTATION

"Nuclear Receptor Signaling Pathways that Regulate Inflammation"
Dr. Christopher Glass, Professor of Medicine and Cellular and Molecular Medicine, School of Medicine, University of California, San Diego.

A current Council member, Dr. Glass has a major research interest in the regulation of macrophage gene expression. His laboratory uses molecular and biochemical approaches to elucidate the mechanisms of action of transcription factors that mediate the biological effects of retinoic acid, vitamin D and other hormone-like molecules during macrophage development and terminal differentiation. These approaches are also used by his laboratory to investigate mechanisms of transcriptional control of macrophage-specific genes that have been implicated in the pathogenesis of atherosclerosis and other diseases. He has served as a member and as chair of the NIH Endocrinology Study Section, and has also chaired or co-chaired a number of Gordon and Keystone conferences. His research has been funded by NIH since 1991. He has published over 125 original articles and has authored 50 invited articles and book chapters. Dr. Glass earned both his M.D. and Ph.D. from the University of California, San Diego.

X. REPORTS OF SUBMCOMMITTEES: CONSIDERATION OF REVIEW OF GRANT APPLICATIONS

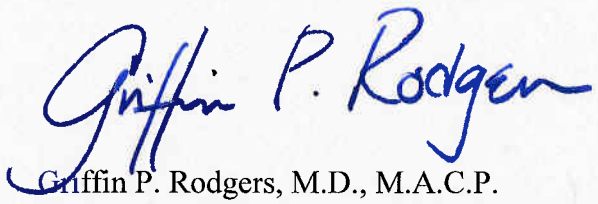
A total of 1,800 grant applications, requesting support of $481,521,097 were reviewed for consideration at the May 12, 2010 meeting. Funding for these applications was recommended at the Scientific Review Group recommended level. Prior to the Advisory Council meeting, an additional 1,126 applications requesting $305,506,283 received second-level review through expedited concurrence. All of the expedited concurrence

applications were recommended for funding at the Scientific Review Group recommended level. The expedited concurrence actions were reported to the full Advisory Council at the May 12, 2010 meeting.

XI. ADJOURNMENT

Dr. Rodgers thanked the Council members and presenters for their attendance and valuable discussion. There being no other business, the 183rd meeting of the NIDDK Advisory Council was adjourned at 4:30 p.m.

I hereby certify that, to the best of my knowledge, the foregoing summary minutes are accurate and complete.



Griffin P. Rodgers, M.D., M.A.C.P.
Director, National Institute of Diabetes and Digestive and Kidney Diseases
Chairman, National Diabetes and Digestive and Kidney Diseases Advisory Council