NATIONAL ACADEMY OF SCIENCES
NATIONAL ACADEMY OF ENGINEERING
INSTITUTE OF MEDICINE
NATIONAL RESEARCH COUNCIL

REPORT IN BRIEF

Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease

A new data network that integrates emerging research on the molecular makeup of diseases with clinical data on individual patients could drive the development of a more accurate classification of disease and ultimately enhance diagnosis and treatment. Recent advances in biomedical research have caused an explosion of data, offering the potential to develop a "New Taxonomy" that defines disease based on underlying molecular and environmental causes, rather than on physical signs and symptoms. This report outlines how research and clinical data can be captured in a "Knowledge Network" that will be broadly accessible to researchers and clinicians. As well as improving health care, the new data network could also improve biomedical research by enabling scientists to access patient information through electronic health records, while still protecting patient rights.

Today, in a clinic somewhere in America, a patient is learning that he has diabetes. Based on the patient's symptoms and lab tests that show high levels of insulin, his doctor diagnoses Type II diabetes—but this imprecise category serves only to distinguish the disease from diabetes that typically occurs at an earlier age (Type I) or during pregnancy (gestational). The doctor will likely prescribe metformin, the most common treatment for Type II diabetes in the United States, but the physician has no way of knowing how the patient will respond to the drug. The doctor can't predict if the patient will experience diabetes-related complications such as kidney failure or blindness, and there are no indicators of the risk of diabetes to the patient's siblings or children.

Consider a world where the scenario is different. Here, biomedical research has revealed a series of molecular and environmental factors that underlie several sub-types of diabetes. Looking at the patient's medical history and genome information along with other molecular data , the doctor is able to identify precisely what the patient is suffering from and devise a customized treatment plan. The doctor can also predict if other family members will likely develop diabetes in the future, and if necessary could implement preventative care.

In recent years, dramatic advances in biomedical research have created an explosion of data that could be used to move toward such a world of improved health outcomes. However, currently there is a disconnect between scientific advances in research and the incorporation of this information in the clinic. It can take years for biomedical research information to reach doctors and patients, and in the meantime wasteful health care expenditures are incurred for treatments that are only effective in specific subgroups. In addition, researchers don't have access to the wealth of clinical data on patients that is collected at the point of care. Overall, opportunities to understand, diagnose, and treat disease more accurately and better inform health care decisions are being missed.

This report provides a framework for developing a more precise and more accurate classification of disease based on molecular biology—a "New Taxonomy" of human diseases—that could revolutionize disease diagnosis, therapy, and clinical decisions, leading to more individualized treatments and improved outcomes for patients.

Benefiting from New Research and Technologies

Advances in research and emerging technologies have the potential to transform medicine. For instance, in recent years the cost of obtaining an individual's genome sequence has fallen rapidly. Genome sequencing could soon cost as little as $1000 and is expected to become a routine clinical test to help doctors better diagnose and treat diseases. The increase in genetic data has already boosted understanding of the root causes behind some diseases and conditions. For example, by looking at the genomes of patients with high cholesterol, scientists found that a significant number of the patients had a non-functional copy of a gene that encodes a low-density-lipoprotein-receptor. For these individuals, lifestyle interventions such as diet and exercise alone were ineffective at reducing the early onset of cardiovascular disease. Identifying these patients would allow their doctors to prescribe statin drugs at an early age rather than first attempting to control cholesterol with diet and exercise. There is strong evidence that the early use of statin drugs in these individuals can provide a therapeutic benefit. The emergence of new technologies to investigate other biological features of disease is expanding abilities to describe and understand disease.

Harnessing Emerging Disease Data

In order to harness the power of emerging disease data, systems are needed to collect and make the information widely accessible. The committee suggested a framework for creating an information system called a Knowledge Network of disease that integrates the rapidly expanding range of information on the causes of disease and allows researchers, health-care providers, and the public to share and update this information (see Figure 1).

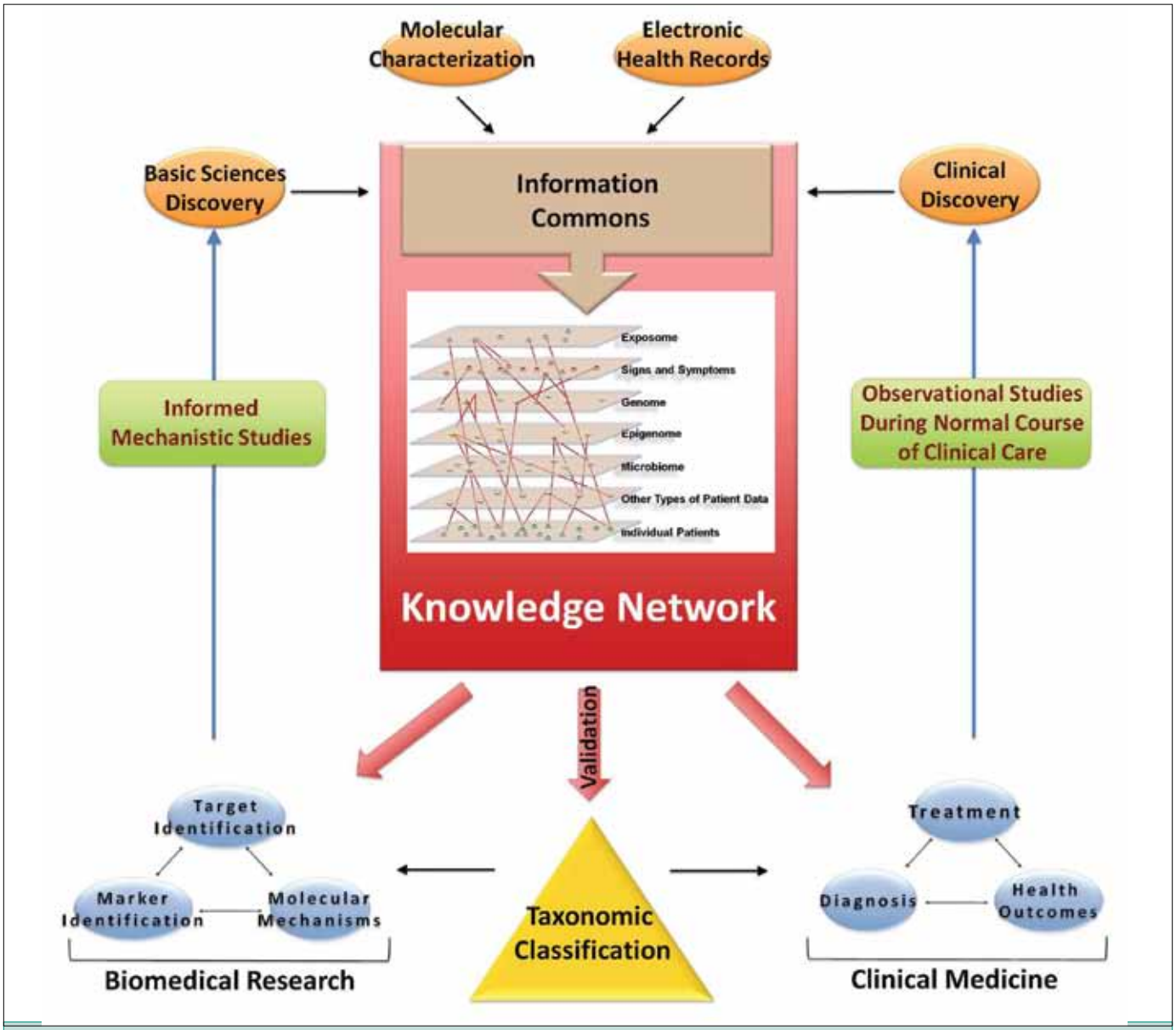

Figure 1. This diagram illustrates a comprehensive biomedical information network that supports a New Taxonomy of disease.

Such a system is centered around an "Information Commons," a data repository that links layers of molecular data, medical histories, including information on social and physical environments, and health outcomes to individual patients (see Figure 2). Data would be continuously contributed by the research community and from the medical records of participating patients.

Using the Knowledge Network

The Knowledge Network would impact all aspects of biomedicine and health care. By analyzing connections between information sets (for example between the genome and environmental exposures) basic scientists would be able to formulate and test disease mechanisms, and clinicians could develop new treatments based on unique features of a disease and tailored to each patient. The availability of more diverse information about each disease would allow insurers and health care providers to more precisely define disease subtypes.

Traditionally, academic research laboratories have had little connection with clinicians' offices. In order to find groups of patients or patient-derived samples for study, researchers have had to use informal referral networks to identify physicians working with patients with diseases of interest. After contributing to the study, the patient is unlikely to remain connected to the research process, or be aware of the outcomes. Because the Information Commons is continuously updated with new information, researchers would be able to conduct long-term follow-ups and better understand how patients respond to treatments over long periods.

The initiative to develop a New Taxonomy—and its underlying Information Commons and Knowledge Network—is a needed modernization of current approaches to integrating different types of data, not an "add-on" to existing research programs. Enormous efforts are already underway to achieve many of the goals of this report, but the committee found that a system-wide emphasis on shifting the acquisition of molecular data to point-of-care settings and the coordination required to ensure research data reach the Information Commons and Knowledge Network is often missing.

How Do We Get There?

The committee outlined a number of steps toward creating an Knowledge Network of disease and deriving a New Taxonomy from it.

- **Conduct pilot studies that begin to populate the Information Commons with data**
  Pilot studies, including those conducted in heath care settings, would help scientists figure out how to integrate molecular data with medical histories and health outcomes in the ordinary course of clinical care. These studies would address the institutional, cultural, and regulatory barriers to widespread sharing of individuals' molecular profiles and health histories while still protecting patients' rights. Much of the initial work necessary to develop the Information Commons should take the form of observational studies, which would collect molecular and other patient data during the normal course of treatment.
- **Integrate data to construct a knowledge network of disease**
  As data from pilot studies begins to populate the Information Commons, substantial effort should go into integrating these data with the results of basic biomedical research in order to create a dynamic,

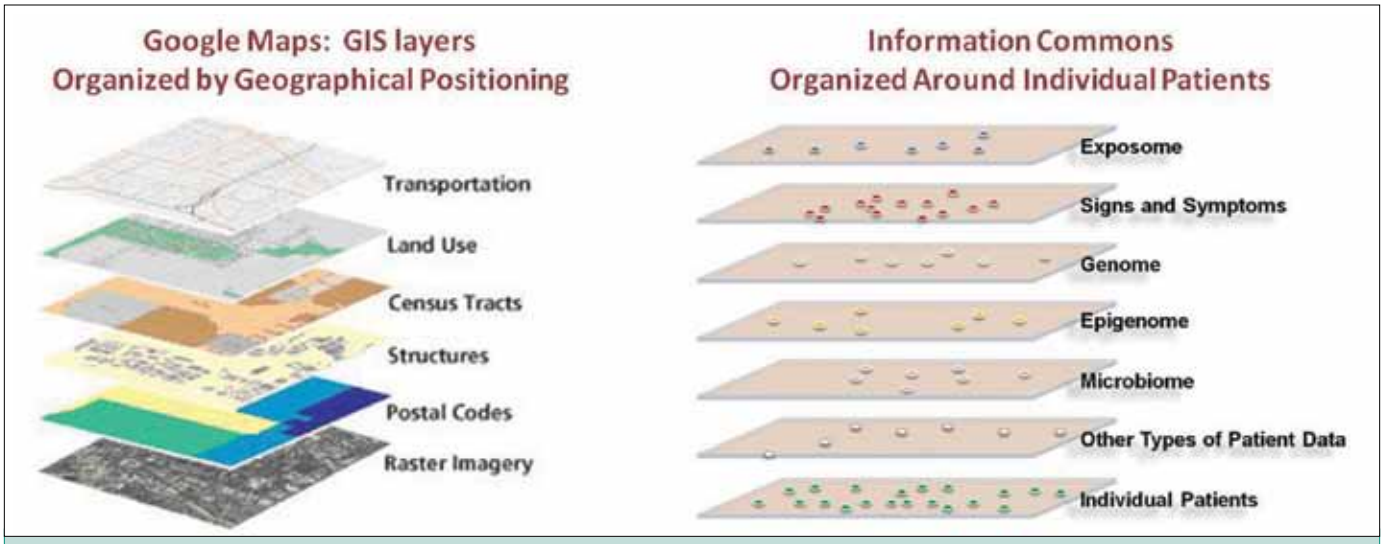

Figure 2. The proposed Information Commons (*right panel*) is somewhat analogous to a Geographical Information System (GIS) such as Google Maps (*left panel*). GIS was launched following public access to Global Positioning System (GPS) data and dramatic improvements in database technology—comparable to current advances in biomedical data generation and handling. Like Google Maps, the Information Commons would consist of multiple layers of data that together provide insights that could not be gained from any of the layers alone. Source: FPA 2011 (*left panel*)

interactive Knowledge Network. This network, and the Information Commons itself, should leverage state-of-the-art information technology to provide multiple views of the data appropriate to the varying needs of different users such as basic researchers, clinicians, or outcomes researchers. The incorporation of electronic medical records into the health care system and the advent of inexpensive ways of collecting health information could also create opportunities to integrate data for the Information Commons more efficiently.

- **Initiate a process within appropriate federal agencies to assess the privacy issues associated with the research required to create the Information Commons**
  Because privacy issues associated with genetic information have been studied extensively, this process need not start from scratch. However, investigators who wish to participate in the pilot studies discussed above—and the Institutional Review Boards who must approve their human-subjects protocols— will need specific guidance on the range of informed-consent processes appropriate for these projects.
- **Ensure data sharing**
  Widespread data sharing is essential to the success of each stage of creating a new disease taxonomy. Most fundamentally, information on how gene sequence translates to symptoms must be broadly accessible so that a wide diversity of researchers can mine them. Data sharing standards that respect individual privacy concerns while enhancing the deposition of data into the Information Commons should be created. These standards should provide incentives that motivate data sharing over the establishment of proprietary databases for commercial intent.
- **Develop an efficient validation process to incorporate information from the disease Knowledge Network into a New Taxonomy of disease**
  Insights into disease classification that emerge from the Information Commons and the derived Knowledge Network will require validation of their reproducibility and their utility for making clinical decisions such as selecting appropriate treatment, before adoption into clinical use. The speed and complexity with which such validated information emerges will undoubtedly accelerate and will require novel decision support systems for use by all stakeholders.
- **Incentivize partnerships**
  A New Taxonomy incorporating molecular data could become self-sustaining by accelerating delivery of better health through more accurate diagnosis and more effective and cost-efficient treatments. However, to cover initial costs associated with collecting and integrating data for the Information Commons, incentives should be developed that encourage public private partnerships involving government, drug developers, regulators, advocacy groups and payers.

Read or purchase this report and locate information on related reports at
http://dels.nas.edu/bls

Committee on a Framework for Developing a New Taxonomy of Disease: Susan Desmond-Hellmann (*Co-Chair*), University of California, San Francisco; **Charles Sawyers** (*Co-Chair*), Memorial Sloan-Kettering Cancer Center; **David R. Cox**, AQG, Pfizer Inc.; **Claire Fraser-Liggett,** University of Maryland, School of Medicine; **Stephen J. Galli,** Stanford University; **David B. Goldstein,** Duke University School of Medicine; **David J. Hunter,** The Harvard School of Public Health; **Isaac S. Kohane,** Harvard Medical School; **Manuel Llinas,** Princeton University; **Bernard Lo,** University of California, San Francisco; **Tom Misteli,** National Cancer Institute; **Sean J. Morrison,** University of Michigan; **David G. Nichols,** The Johns Hopkins University School of Medicine; **Maynard V. Olson,** University of Washington; **Charmaine D. Royal,** Duke University; **Keith R. Yamamoto,** University of California, San Francisco; **India Hook-Barnard** (*Study Director*), **Amanda Mazzawi** (*Senior Program Assistant,* Oct. 2010–May 2011), **Pagnola Golson** (*Senior Program Assistant,* June 2011), **Orin Luke** (*Senior Program Assistant,* July–Sept. 2011), **Melinda Divito** (*Christine Mirzayan Fellow,* Oct.–Dec. 2010), National Research Council.

The National Academies appointed the above committee of experts to address the specific task requested by the National Institutes of Health. The members volunteered their time for this activity; their report is peer-reviewed and the final product signed off by both the committee members and the National Academies. This report brief was prepared by the National Research Council based on the committee's report.

For more information, contact the Board on Life Sciences at (202) 334-1534 or visit http://dels.nas.edu/bls. Copies of *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* are available from the National Academies Press, 500 Fifth Street, NW, Washington, D.C. 20001; (800) 624-6242; www.nap.edu.