Blue Ribbon Panel
Future Strategies for Bioinformatics in CDC's Infectious Diseases Laboratories
June 8-9, 2011

INTERNAL REPORT

On June 8-9, 2011, CDC's Office of Infectious Diseases (OID) convened a panel of external consultants with expertise in the fields of bioinformatics, informatics, and/or laboratory information technology to

- Review the current status of bioinformatics resources across the infectious diseases laboratories at CDC
- Identify critical gaps in CDC's bioinformatics resources as compared with other similarly purposed programs
- Provide input to CDC on the best strategies to develop and maintain the bioinformatics personnel and infrastructure needed to sustain the laboratory and epidemiologic science that is expected of the agency by policy makers, the scientific community, and the public.

CDC program leaders and laboratorians provided data, information, and perspectives to the panel through pre-read materials, presentations, and dialogue with panel members.

Background

- The life sciences have undergone a technological revolution over the last decade. With advances in genomic methods, such as second- and third- generation high-throughput sequencing, it now possible to generate massive amounts of data within a very short period of time. For public health in general, and CDC in particular, this capacity is revolutionizing the approach to multiple activities such as core infectious disease surveillance, outbreak management, vaccine development, strain subtyping, development of laboratory diagnostics, and detection of drug resistance.
- The increasing ease and decreasing cost of high-throughput technologies are on the verge of making microbial genomics, metabolomics, proteomics, etc., a standard component of laboratory science. The challenge is no longer doing the sequencing but having the computational infrastructure needed to manage, analyze, and interpret the data.

Observations of the Panel

- There does not appear to be a coherent strategic vision of how genomics and bioinformatics should be employed to leverage the core public health mission of CDC. In particular, CDC

needs to recognize that bioinformatics is not simply a support mechanism; the science of bioinformatics has emerged as a distinct and independent research discipline that needs to have equal standing with epidemiologic and laboratory science for public health.

- Compared with other major institutions in the United States that conduct public health-related research and application, CDC seems ill-prepared to accomplish the rapid, data-intensive computations needed for disease outbreak situations, except to a limited degree in specific circumstances (such as the H1N1 pandemic). In other recent high-profile situations (such as the cholera epidemic in Haiti), CDC has trailed the rest of the field. This is a visible reflection of CDC's capacities, and it has also impacted public health messaging.
- CDC lacks the organization, expertise, and computational infrastructure needed for data management and analysis of the genomic information it is now capable of acquiring, much less the logarithmic increase in the amount of data that will need to be integrated from multiple and diverse platforms, such as next-generation sequencing, metabolomics, and proteomics, in the near future. The excellent but limited-scale high-performance computing group is a promising start, but the group is relatively siloed and most of the funding is short-term.
- Many genomics/bioinformatics projects are being carried out through collaborations with outside institutions, but several CDC groups lack the expertise to identify external groups with which to engage.
- Analysis and appropriate interpretation of data has been one of CDC's traditional strengths, but CDC runs the risk of going from outdated to obsolete, and then to irrelevant if the Agency does not recognize the need to invest in the science of bioinformatics.

Recommendations

In developing a vision for CDC's bioinformatics program, the panel proposed short-term goals (1-3 years) to meet the most urgent needs and opportunities and longer-term goals (5 years) that anticipate the scientific developments that will be occurring in order to ensure that genomics and bioinformatics capabilities at CDC become sustainable resources for public health.

Short-term goals

- Create an environment that
  - Recognizes that bioinformatics is a third public health scientific discipline alongside epidemiology and laboratory science
  - Recognizes that investing in the science of bioinformatics is essential to keeping pace with the revolution in the life sciences
  - Recognizes that the science of bioinformatics is an integral part of CDC's public health mission and must be considered when overall strategies and funding priorities are determined

  - Recognizes the difference between scientific and business computing needs and facilitates scientific computing and analysis.
- Establish a core group of bioinformatics scientists, managed and led by a respected bioinformatics specialist, to meet the needs for analysis and interpretation of public health data generated using high-throughput technologies.
  - Adopt the concept of having bioinformatics scientists embedded within individual programs and ensure that those scientists are linked with the core group. This model is in use at many other institutions.
  - The bioinformatics core group should have training as a central part of their mission. The core group should be charged with keeping up with current software tools and approaches that are most relevant to genomic research and the public health mission of CDC. This should be done in close collaboration with individual programs and the embedded bioinformatics scientists.
- Given how inexpensive sequencing has become, it will be critical for CDC to determine the extent to which basic genomics and bioinformatics services should be provided in-house versus outsourced. Partnerships do not necessarily require large funding inputs.
- Leverage other resources by fostering collaborations with genomics and bioinformatics groups in other government agencies, academia, etc., to access the latest expertise. Such collaborations could become a major avenue by which CDC bioinformatics expertise and software resources are strengthened. Collaborations can also provide CDC with working models for how to build bioinformatics infrastructure.
- Take advantage of the substantial bioinformatics expertise available at local academic institutions. Many of these have researchers who would willingly work with CDC on public health problems. Establish a formal training and/or intern program; students in local institutions are a tremendous potential resource for CDC.
- Work with local academic institutions to establish public health bioinformatics programs to train the current and future generations of bioinformatics scientists for CDC and state health departments.
- Provide leadership to the state public health departments on using genomics, other "omics," and bioinformatics to meet core public health missions. Short of this, states will start working more and more with their local academic institutions who do have these capabilities, precluding CDC involvement in this increasingly important area of public health.

Longer-term goals

Ensure that

- CDC is recognized as a leader and active participant in the use of "omics" and bioinformatics for public health, and remains a credible source of public health information.
- CDC is using a balance of internal resources and outside collaborations to meet core surveillance and outbreak response needs and to keep up with changing technologies.

- CDC has created an environment that understands and supports scientific computing and recognizes bioinformatics as both an essential support function and also a scientific partner in achieving its public health mission.
- CDC has helped developed the courses and programs needed to train and develop the bioinformatics human resources to meet the growing need in public health.

Conclusion

- The authoritative analysis and interpretation of scientific data is a critical area where CDC must excel for the Agency to continue to provide international leadership in public health and policy. The mission of CDC will be seriously compromised if the organization does not have the appropriate bioinformatics expertise and computational infrastructure.
- To avoid becoming outmoded and to maintain relevance, CDC needs to make substantial and sustained commitment to developing the bioinformatics personnel, computing resources, and training needed to perform the data management, analysis, and interpretation expected from the nation's leader in public health.
- Unless severe shortcomings in bioinformatics expertise and infrastructure are overcome, efforts to adapt high-throughput technologies to infectious diseases surveillance, outbreak management, vaccine development, and other critical components of CDC's public health mission will not succeed. Should this occur, CDC will be at risk of going from outdated, to obsolete, and then to irrelevant.