After four years, the seven National Centers for Biomedical Computing (NCBCs)—established largely to build a national biocomputing infrastructure—have, as one might expect, produced an impressive array of computer tools. >

NCBC Update: Shedding New Light On BIOLOGICAL COMPLEXITY

By Katharine Miller

But it's the Centers' wide-ranging impact on biomedicine that takes center stage. From AIDS to diabetes, prostate cancer or schizophrenia, the NCBCs are changing the landscape of disease research by shedding new light on biological complexity.

"The impact on biology and medicine happened faster than anyone expected," says **Russ Altman, MD, PhD**, co-principal investigator for **Simbios, the National Center for Physics-based Simulation of Biological Structures**, an NCBC grantee at Stanford University.

And that impact springs from the way the NCBCs function, says **Andrea Califano, PhD**, who heads the **National Center for Multiscale Analysis of Genomic and Cellular Networks (MAGNet)** at Columbia University. "Developing new tools in the context of solving specific scientific, biological or medical problems is what I think has allowed the NCBCs to successfully penetrate the broader community with tools, techniques and methodology," he says. "We've shown what can be accomplished by applying these tools to biological problems."

And while the specific breakthroughs enabled by NCBC tools varies with the tool being used or the disease being studied, it is clear that they are all helping researchers approach the complex system that is the human body. "Dealing with complexity is the essential challenge of this century in biology," says **Scott Delp, PhD**, co-PI for Simbios. "And you can't do it without computers."

"Developing new tools in the context of solving specific scientific, bicalifano, biological or medical problems is what I think has allowed the NCBCs to successfully penetrate the broader community with tools, techniques and methodology." says Andrea Califano.

NCBC 10 YEAR GOALS

Califano: **"One critical thing we hope to accomplish is to create a new breed of biologist trained both in computational and experimental sciences. You already see evidence of this in some labs. Now, as never before, some of the projects enabled by the NCBCs have computation and experimental biology playing hand in hand rather than in a pipeline fashion. That is also reflected in the tools that we generate. Unlike other platforms, geWorkbench was created for an experimental biologist who wants to learn enough computational biology to be able to analyze data. It's easy and intuitive to use, and the researcher doesn't have to learn complex scripting languages. The emphasis has been on enabling experimental labs to use more and more computational tools. Across the entire set of activities at MAGNet, the real aim is to fuse the two disciplines and to create a really interdigitated boundary between the computational and experimental life sciences."**

Andrea Califano, PhD, is the principal investigator for the National Center for Multiscale Analysis of Genomic and Cellular Genomics (MAGNet) and professor of biomedical informatics at Columbia University.

MAGNet

Here, following a few years of hard work, the NCBC PIs reflect on what they've accomplished so far, how they've gained traction in the research community, and what their goals are going forward.

NCBC Tools: Enabling Discovery Across The Disease Spectrum

From the start, each NCBC's tool and infrastructure development goals were driven by a cluster of specific biological problems—commonly referred to in NCBC parlance as the "driving biological problems" or DBPs. After a few years, these DBPs were replaced by a new set of DBPs, ensuring that the tools would be suitable for multiple purposes. That strategy has worked.

"To a certain degree, the tools and biology are push-pull kinds of associations," says **Art Toga, PhD**, principal investigator for the **Center for Computational Biology (CCB)** based at the University of California, Los Angeles. "The tools get developed because you couldn't do something without them. And vice versa, you get this tool and you decide to pose new questions. You end up pushing and pulling so that both are advanced."

Thus, NCBC tools that were developed to address one biomedical problem have proven to be broadly useful. For example, at **i2b2—Informatics for Integrating Biology and the Bedside**—an NCBC based at Harvard, tools developed to allow the use of medical record systems for clinical research initially focused on diseases such as asthma, obesity and depression. Now, however, these tools have been adopted at 18 large academic health centers with no apparent limit on the number of diseases that can be studied, says i2b2 principal investigator **Zak Kohane, MD, PhD**.

And imaging tools originally developed by CCB and the **National Alliance for Medical Image Computing (NA-MIC)** to study schizophrenia in the brain are now proving useful in studying many other brain diseases, as well in prostate cancer (at NA-MIC) and cardiovascular disease (in CCB's case). "You can begin to see how the shape-modeling approach [we've developed] is applicable to a whole range of biological problems," says Toga of CCB.

Similarly, OpenSim, a software program developed by Simbios to study human movement and movement disorders, was first used to conduct research into one of the Simbios DBPs, cerebral palsy, but is now being used more broadly. Indeed, it has been adopted by more than one thousand individuals working on any number of problems including osteoarthritis, Parkinson's disease and stroke.

This is the vision of the NCBCs—to provide the underlying computational tools that will advance the field of medicine and biology, across a spectrum of diseases. As **Mark Musen, PhD**, of the **National Center for Biomedical Ontologies (NCBO)** at Stanford, says, "We're enablers. We are providing the foundation by which

"Dealing with complexity is the essential challenge of this century in biology," says Scott Delp. "And you can't do it without computers."

Kikinis: **"We are developing algorithms and a platform—the NAMIC kit—for analysis of diagnostic images. I think that platform will be one of our major accomplishments. It is free and open source with a very liberal license, and it will continue to be developed. That will continue for a long time. So our goal is to develop enabling technologies and make them accessible. That will be one of the legacies of the center."**

NCBC 10 YEAR GOALS

NA MIC

Ron Kikinis, PhD, is the principal investigator for the National Center for Medical Image Computing (NA-MIC) as well as director of the Surgical Planning Laboratory of the Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, and professor of radiology at Harvard Medical School.

investigators can do research that will impact human health. Our goal is to create the kinds of tools that would be valuable to everybody."

Ron Kikinis, PhD, head of NA-MIC, concurs. "We will not solve cancer but we will provide the people who are fighting cancer with better tools to fight their fight," he says. "And the DBPs will use these tools and promote those tools into their communities—so that makes it possible for lots of different diseases to be addressed."

Brian Athey, PhD, co-PI for the National Center for Integrative Biomedical Informatics, centered at the University of Michigan, agrees. While his center's tools have contributed to a better understanding of type 2 diabetes and prostate cancer progression, the tools' reach extends much farther: "We're opening doors to new research," he says.

"The tools get developed because you couldn't do something without them. And vice versa, you get this tool and you decide to pose new questions. You end up pushing and pulling so that both are advanced," Art Toga says.

NCBC Challenge: Putting It All Together

For the last thirty years, biology has been about breaking things down into their fundamental parts to understand them. "But things don't work as independent parts," says Delp. "Theoretical and computational biology let you put things back together to understand the whole system."

Several of the NCBC PIs cite the reassembling of biological pieces as a major focus of their efforts. For example, literally thousands of experiments have looked at how elements of the neuromuscular system (muscles, joints, connective tissue) operate independently. But, Delp says, looking at those elements separately doesn't tell you how people move. OpenSim lets researchers put the pieces together. "When you can code the details accurately in a computer framework, then you can understand how the system works," Delp says.

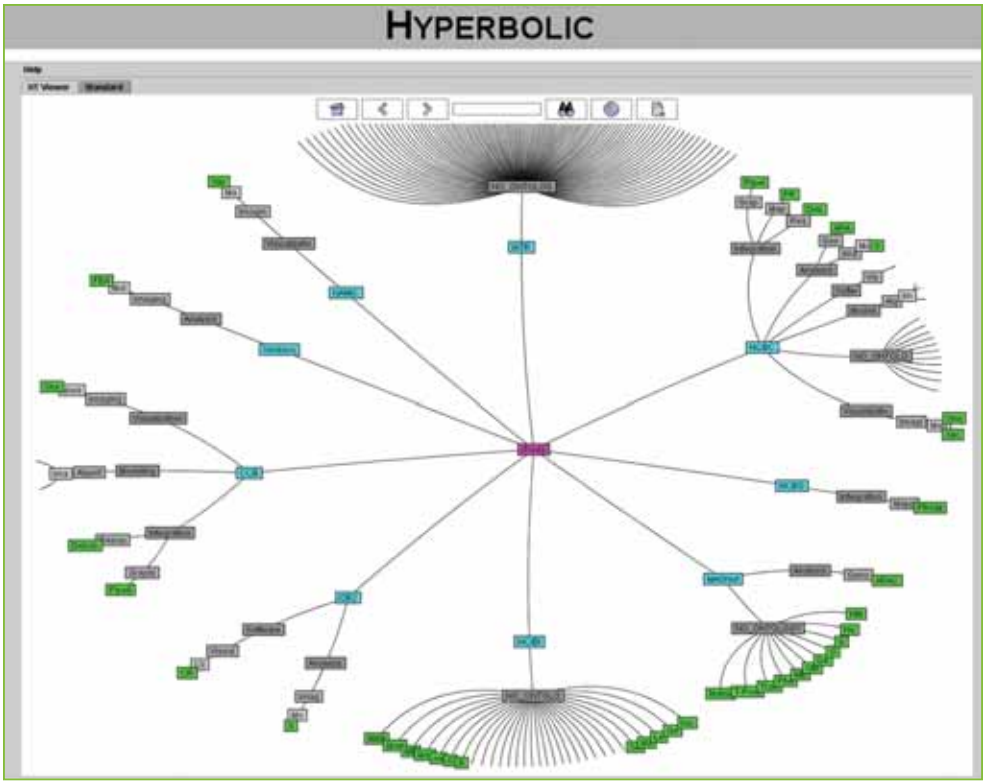
Likewise for the brain, says CCB's Toga. Brain researchers have typically focused on only one variable at a time—for example, electrical activity, blood flow, distribution of receptors, gene expression patterns, or cortex morphology. But, Toga says. "All of these brain changes are happening in concert." To understand the brain requires re-integration of these events. CCB, Toga says, is providing the tools, mechanisms, and strategies to put things

Toga: **"Our hope is to continue to integrate what we know about the brain in a way that allows us to ask questions such as: 'How does the brain change throughout a person's life?' These sorts of emerging questions are provocative. And we can only ask them because of computation. So by the end of our ten years, we really hope that our center produces new research programs that can continue to evolve in accord with the basic thrust of the NCBCs. Because you know, it doesn't finish. They haven't finished mapping the earth yet and there's only one of those! How can anyone possibly suggest we will ever finish mapping the human brain when there are billions of them? So we'll continue to layer on what we already know without throwing away our previous efforts."**

NCBC 10 YEAR GOALS

Center for Computational Biology (CCB)

Arthur Toga, PhD, is the principal investigator for the Center for Computational Biology (CCB), and a professor of Neurology and Director of the Laboratory of Neuro Imaging at the University of California, Los Angeles.



Working together NCBC researchers created iTools—a way to manage the description of computational biology data, tools, and services. Using the iTools hyperbolic viewer a researcher can displays all of the activities of the NCBCs organized by Center (as shown here) or by activity. iTools also lays the groundwork for interoperability among diverse biomedical computing tools. Reprinted from Dinov, ID, et al., 2008 iTools: A Framework for Classification, Categorization and Integration of Computational Biology Resources. **PLoS ONE *(2008) 3:(5):e2265.***

back together. "Observations from one project in 2007 can be combined with other observations in another laboratory using different subjects and techniques in 2008," Toga says. "That transition in science is revolutionary, and the computational strategies that enable it are only now beginning to emerge."

MAGNet hopes to provide a similar service at the genetic and cellular level. Very few diseases are caused by a single gene, Califano says. Usually a complex interplay of genetic and epigenetic factors is involved. "But what has been lacking is a framework for integrating genetic,

"We're enablers." Mark Musen says. "We are providing the foundation by which investigators can do research that will impact human health. Our goal is to create the kinds of tools that would be valuable to everybody."

Musen: **"We are thinking about what it would mean to be able to move biomedical knowledge from prose to machine-processable format. The long-term vision is to create the infrastructure and tools so that biomedical literature could be intelligible to both people and machines. Ultimately this could allow intelligent computer-based agents to read the literature, to make associations between scientific contributions, and to synthesize ideas from the literature. That would obviously change the way we do science in a very profound way. But there are lots of baby steps until we can do that."**

NCBC 10 YEAR GOALS

Mark Musen, MD, PhD, is the principal investigator for the National Center for Biomedical Ontologies (NCBO) and professor of medicine at Stanford University School of Medicine.

NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

epigenetic, functional and structural data—and getting an answer that can really dissect disease," he says. MAGNet's goal is to establish such a framework and to show that the framework can integrate data in meaningful ways for several diseases. "We already have proof of concept for glioblastoma multiforme—a cancer that produces the worst possible prognosis in patients," Califano says. The results for that work will be published in the next few months. "This kind of proof of concept in a disease is of course important, but at the same time the methodology becomes universal."

NCBCs: More than the sum of their parts

The NCBCs are also working together in various ways to ensure that they have a broad impact. In some ways this is a surprise, say the NCBC PIs, because the NIH cast such a wide net—with centers that cover ontologies, simulations, clinical systems, systems biology and imaging. "Given the breadth of the needs and the solutions to biomedical computing problems," says Kohane, "it wouldn't have been surprising if there had been no overlap and the synergies had been fewer."

"What bioinformatics was five years ago is frankly just a glimmer of what it is today," Brian Athey says. "It's exploding into something much more robust. And that's going to continue for a while."

NCIBI is also integrating many different high-throughput data types to better understand complexity. "We do not yet understood the full complexity of the architecture of the human genome," Athey says. "Only 2 percent of the genome are 'genes' and we're learning more and more that the other 98 percent are doing things." To tackle that problem, he says, computational biology is making huge strides. "What bioinformatics was five years ago is frankly just a glimmer of what it is today," Athey says. "It's exploding into something much more robust. And that's going to continue for a while."

Yet the NCBCs have found overlap and have helped each other. For example, the i2b2 center collaborated with NCIBI around Type 2 diabetes, Kohane says. And NA-MIC nicely complemented i2b2's major depression DBP by correlating patient imaging with what was being seen genetically. Similarly, ontologies from NCBO have been helpful to CCB in constructing their brain atlas; and CCB and Simbios have used some of NA-MIC's visualization tools.

Even though the NCBCs might be developing different tools, Califano says, "when you

Athey: "There's much more work to do to figure out how to use systems biology more effectively to understand disease and its complications. The daunting complexity of biological systems is becoming more and more clear. To gain an understanding of that complexity, we need an integrative approach that's iterative and that allows the integration of many different kinds of data types around hypotheses and models. The abundance of high throughput data we're presented with from next generation sequencing, and what that's revealing about the transcriptome and alternative splicing, and all the components we haven't yet annotated—it's just astounding. It's literally changing our basic understanding of cells and their complexity and function. And, frankly, it's changing what our understanding of a gene is. So there's a lot of work to do. I think that's the theme. And each success brings on new challenges."

Brian Athey, PhD, is the principal investigator for the National Center for Integrative Biomedical Informatics (NCIBI), associate professor of biomedical informatics at the University of Michigan, and director of the Michigan Center for Biological Information.

NCBC 10 YEAR GOALS

NCIBI
National Center for Integrative Biomedical Informatics

tackle a biological problem you must tackle it from several angles." So for example, MAGNet and NCIBI have several DBPs that focus on analyzing genomic data as a way of studying neurodegenerative diseases, diabetes or cancer. But these same diseases also need to be studied using data from large cohorts, which ties in to what i2b2 does at Harvard to use medical records to study large populations. It also ties in to the ontology work of Mark Musen, Califano says, because ontologies provide an essential foundation for other work. And, he says, when you look at the actual problem you're trying to understand, all sorts of issues related to physical modeling also come up. Indeed, according to Altman, eventually cellular physics will become an essential piece of systems biology.

"The reality of why all the centers come together is precisely around the biology, Califano says. "We develop all the different techniques and infrastructure to tackle biology problems, but when you actually want to tackle one of these problems, you require all of these approaches."

And those multiple tools also need to be kept organized. So one key activity that has united all the centers, says Musen, is the creation of an online tool that allows biomedical software resources to be easily identified and searched online. Called Biositemaps, the tool, seeded with information about the NCBC tools, can inform search engines about software available from any organization that creates a simple Biositemap file as described on the site (http://www.biositemaps.org). NCBO is providing the ontology behind the tool but, Musen says, "It's a product of all the NCBCs that would not have been possible without the cooperative involvement of all the different centers."

"If we actually successfully did a big population study and discovered something important or successfully calculated how to design a vaccine or predicted a new drug for a specific disease, then we'd be bringing ourselves to the next level," says Kohane. "We'd be solving a biomedical problem of true health relevance. In fairness, I think we're all trying to get there, but we're not there yet."

NCBC 10 YEAR GOALS

Kohane: "Within the ten-year time frame, the goal would be to establish a kind of scientific ecosystem around the country where we can use entire healthcare systems as a unit of study. We'll be able to look at reproducibility across multiple academic health centers to see if we're seeing, for example, the same adverse drug events (so that we can push early warnings to prevent such events); or compare efficacious therapies; or compare whether we have reproducible findings in genomics or proteomics across populations. This approach will allow us to do research in a more cost-effective way. And although it sounds venal to talk about cost, cost is a key rate-limiting factor in large population studies. So if we can do clinical research, including genomic measurements in populations of 10,000 to 100,000, that's really a game-changer."

Isaac Kohane, MD, PhD, is the principal investigator for Informatics for Integrating Biology and the Bedside (i2b2), as well as Lawrence J. Henderson Associate Professor of Pediatrics and Health Sciences and Technology at Harvard Medical School, and Chair of the Informatics Program at Children's Hospital, Boston.

i2b2

NCBCs: Bench to bedside

Whether casting a wide net to enable research in lots of areas is enough to render the NCBCs successful remains to be seen. Curing a disease would be better. "If we actually successfully did a big population study and discovered something important or successfully calculated how to design a vaccine or predicted a new drug for a specific disease, then we'd be bringing ourselves to the next level," says Kohane. "We'd be solving a biomedical problem of true health relevance. In fairness, I think we're all trying to get there, but we're not there yet."

"The challenge is," says Delp, "that it takes time to build the tool, teach people how to use it, get it adopted, make a discovery and then translate that into clinical care." Currently, says Delp, "OpenSim is only halfway down that pipeline and is just beginning to see the first examples where new discoveries will enhance human health."

"Adoption by companies is one indication that what we're doing will eventually make a difference to clinical practice," Ron Kikinis says. "We are not yet at that point, but I have these early indicators."

Kikinis says NA-MIC's tool kit is similarly poised for bedside use. He's beginning to see the first signs—such as questions at seminars, and email inquiries—that companies are interested in it. "Adoption by companies is one indication that what we're doing will eventually make a difference to clinical practice," he says. "We are not yet at that point, but I have these early indicators."

Migrating computational biology from the bench to the bedside remains a challenging goal for all the centers. But, as Toga sees it, "I think these computational strategies, which are the hallmark of this program, are having a great effect on accelerating that." CCB is modeling the effect that HIV and Alzheimers have on the brain. These are diseases that will strike people we all know, Toga notes. "So our work immediately transforms a mathematical problem [shape modeling] into something with obvious and immediate clinical value," he says. "And the time frame for doing that is getting shorter and shorter and shorter."

Kikinis summed it up succinctly: "What are the NCBCs doing for biology? Everything. That's by design, but now you can say that they're actually delivering, and there's a sense of excitement. It's clear that things are moving." □

Delp: "The goal is twofold, really. One, that we'll produce a set of tools that are ubiquitous in biomedical research so that every investigator who is interested in how physics affects biological function will have SimTK-based tools as part of their laboratory. The second objective is that we and others will use those tools to make new discoveries that enhance human health."

NCBC 10 YEAR GOALS

simbios

Scott Delp is co-principal investigator for the National Center for Physics Based Simulation of Biological Structures (Simbios) and a professor of bioengineering and mechanical engineering at Stanford University.