DIVERSE DISCIPLINES, ONE COMMUNITY

Biomedical Computation REVIEW

Published by Simbios, an NIH National Center for Biomedical Computing

Mobilizing BIG DATA to Understand Mobility

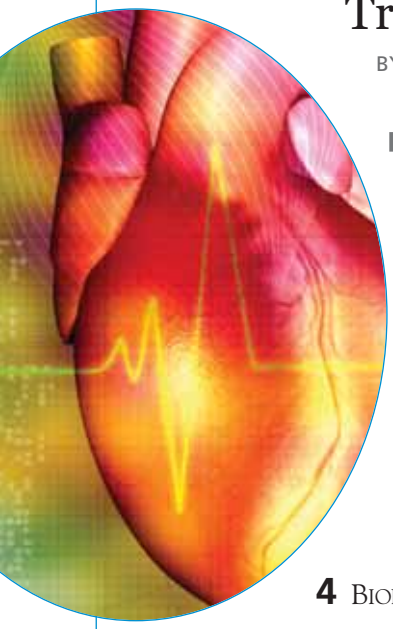
PLUS:

DOING THE HEART GOOD: Translating Models to the Clinic

Spring 2014

contents

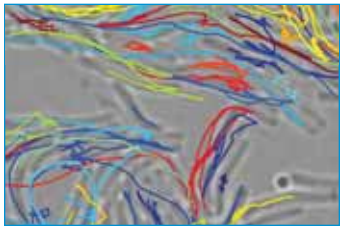
ContentsSpring 2014

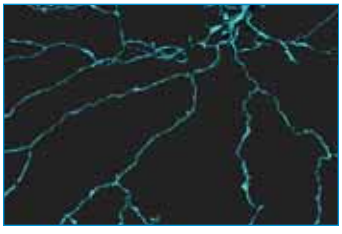


DEPARTMENTS

Cover and Page 16 Art: Created by Rachel Jones of Wink Design Studio using: Simulated skeleton, courtesy of Jennifer Hicks of OpenSim; Man image, © Woodooart, and Binary tunnel image, © Frenta, both from Dreamstime.com.

Page 20 Art: © Krishnacreations | Dreamstime.com.





Spring 2014
Volume 10, Issue 1
ISSN 1557-3192

Executive Editor Russ Altman, MD, PhD

Advisory Editor David Paik, PhD

Associate Editor Joy Ku, PhD

Managing Editor Katharine Miller

Science Writers
Alexander Gelfand, Esther Landuis, Katharine Miller, Kristin Sainani, Stephanie Sides, Sarah C.P. Williams

Community Contributors
Jeanette Schmidt, PhD, Guy Haskin Fernald

Layout and Design
Wink Design Studio

Printing
Advanced Printing

For general inquiries, subscriptions, or letters to the editor,
visit our website at
www.biomedicalcomputationreview.org

Office
Biomedical Computation Review
Stanford University
318 Campus Drive
Clark Center Room S271
Stanford, CA 94305-5444

Biomedical Computation Review
is published by:

simbios
The NIH National Center for Physics-Based Simulation of Biological Structures

Publication is made possible through the NIH Roadmap for Medical Research Grant U54 GM072970. Information on the National Centers for Biomedical Computing can be obtained from http://nihroadmap.nih.gov/bioinformatics. The NIH program and science officers for Simbios are:

Peter Lyster, PhD (NIGMS)
Grace Peng, PhD (NIBIB)
Jim Gnadt, PhD (NINDS)
Peter Highnam, PhD (NCRR)
Jennie Larkin, PhD (NHLBI)
Jerry Li, MD, PhD (NIGMS)
Nancy Shinowara , PhD (NICHD)
David Thomassen, PhD (DOE)
Janna Wehrle, PhD (NIGMS)
Jane Ye, PhD (NLM)

GuestEditorial

BY JEANETTE SCHMIDT, PhD, VICE PRESIDENT FOR INFORMATICS AT AFFYMETRIX

The Uncertain Future of a Magazine

Ten years ago, when **Scott Delp, PhD**, and **Russ Altman, MD, PhD**, decided to write a grant for a National Center for Biomedical Computing (NCBC), their brainstorming sessions with then-postdoc **David Paik, PhD**, sprouted an audacious idea: using part of the grant's dissemination core to create a magazine for the field of computational biology. When the award came through—creating Simbios, the National Center for Physics-Based Simulation of Biological Structures, located at Stanford University—that inspiration became this magazine, *Biomedical Computation Review* (*BCR*).

Under the steady hand of Paik, who is now assistant professor of radiology at Stanford University, the magazine launched in the summer of 2005 with the tagline: Diverse disciplines, one community. The ambitious goal: to be a unifying force for a field that takes a computational slice through the entire realm of biomedicine.

Since that time, Simbios has published an issue of *BCR* every three or four months, covering every conceivable computational topic. Articles have included stories on such computation-heavy fields as genomics, epigenomics, structural biology, computational biomechanics, agent-based modeling, data mining, the physiome, integrative analysis, big data analytics, imaging and connectomics. In addition, there have been health-focused stories about the use of computation to study cancer, aging, Alzheimer's disease, influenza, tuberculosis, cardiovascular diseases, and HIV/AIDS. *BCR* has also covered what I would call the business of computational biology, from educational programs, to funding for the field, computational startups, and the nature of interdisciplinary collaboration. There have also been profiles of women in the field and of computer scientists who made the leap into computational biology and biomedicine. And several of the most read feature stories addressed the field's key challenges: dealing with skeptics and validating models.

In other words, the magazine has delivered on its tag line—providing a sense of identity for a dispersed and diverse community.

Now, with Simbios facing the end of funding for the NCBC program this summer, *BCR*'s future is uncertain, which makes this a great time to reflect on the magazine's value.

BCR has covered a breadth of topics far beyond the scope of the Simbios center, and has kindled many discussions and collaborations. It has been used in the high school classroom to introduce computational topics, has provided an opportunity for students and researchers to both read about interesting topics as well as showcase their work in a less formal context than a traditional journal. It has been a travel companion, providing interesting stories, and has offered a convenient way for funding agencies to show some of the value of their centers of excellence. It has been a pleasure to see it on the magazine racks of companies as well as those of the National Institutes of Health.

I hope that *BCR* will continue in some form. It has been a wonderful addition to the field, has provided educational and fun reading and I would hate to refer to it in past tense. □

SimbiosNews

BY JOY P. KU, PhD, DIRECTOR OF SIMBIOS

Putting Exacycles and Markov State Models to Work on GPCRs

Despite being well-studied, much remains unknown about the dynamics of G-protein coupled receptors (GPCRs), molecules that are prominent drug targets. Recent work published in the journal *Nature Chemistry* breaks new ground in both our understanding of GPCRs and in methodologies for simulating such molecules. Using Google Exacycle and Markov State Models, the study by Google research scientist **Kai Kohlhoff, PhD**, and Simbios collaborators **Vijay Pande, PhD**, and **Diwakar Shukla, PhD**, achieved an unprecedented and insightful millisecond simulation of the GPCR beta-2 adrenergic receptor (β_2AR).

"The impact of this paper is not just that we matched what others found experimentally," says Kohlhoff. "We've gone beyond that and shown the activation mechanism of a GPCR."

The achievement was made possible by combining Google hardware with software from Simbios researchers. Normally, simulating a millisecond of a reaction involving a large molecule like β_2AR would require millions of days on a fast computer or access to a specialized resource such as Anton, the supercomputer designed specifically for molecular dynamics (MD) simulations. But the collaboration with Kohlhoff, a previous Simbios postdoctoral fellow, offered a different solution: tens of thousands of shorter independent simulations of β_2AR on Google Exacycle, a cloud computing infrastructure that transforms Google's spare computing cycles into what is known as an "embarrassingly parallel" system, where there is minimal communication between individual computers. The resulting simulations were then assembled into a single model using Markov State Models (MSMs) to capture 2.15 ms of β_2AR dynamics.

"Cloud resources are much more accessible to the general scientific community [than specialized hardware], and I think that we've shown here that, with the right method and algorithms, you can do the same quality of work," Pande said in an interview with the Stanford News Service.

For its part, Google is pleased with its investment. This study was the first to participate in the Google Exacycle for Visiting Faculty program. "It has shown that the cloud can be used as a new research tool and is worth the time to investigate," says Kohlhoff.

With Google Exacycle, the team generated many trajectories first, exploring them later for insights—a shift from the conventional approach of setting up a simulation to prove a predefined hypothesis. "Traditional simulations are often very narrowly defined and might miss important information," says Kohlhoff. "Our approach is better suited for exploratory research with a lower risk of introducing human bias from the start."

Google's data center, courtesy of Kai Kohlhoff, Google.

From all that data, they generated the first molecular-level description of GPCR activation pathways and identified a large number of previously unknown states that could help in the design of more selective drugs, potentially causing fewer side effects.

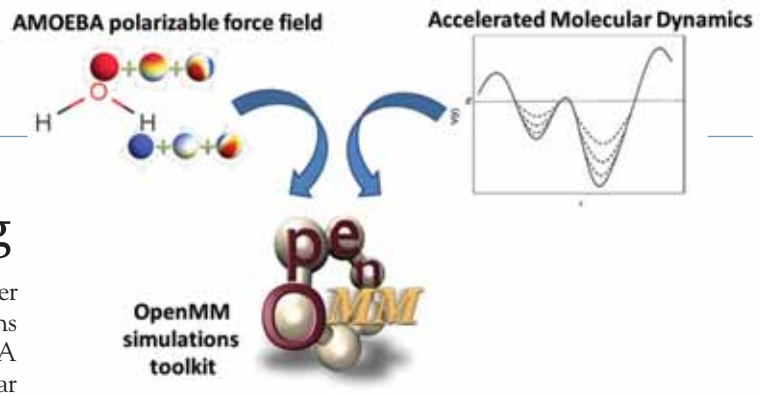
"We only have a few structures capturing the inactive and active states of GPCRs because it is very challenging to crystallize them," says Shukla. "However, as we've shown with this study, with computer simulations, we can even identify infrequently visited intermediate states. That's one big scientific plus."

Kohlhoff has high hopes for the study. Beyond the impact for GPCR research, Kohlhoff says, "We hope this study gets other people to rethink the way they do science." □

DETAILS

To learn more, read the full publication: Kohlhoff, KJ, *et al.*, "Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways," 2014, *Nature Chemistry*, 6:15-21. The MSMBuilder software used to produce the Markov State Models is available at http://simtk.org/home/msmbuilder, and links to the resulting GPCR simulation results will be accessible through https://simtk.org/home/natchemgpcrdata.

Simbios (http://simbios.stanford.edu) is the National Center for Physics-Based Simulation of Biological Structures at Stanford.



Reprinted with permission from Steffen Lindert, Denis Bucher, Peter Eastman, Vijay Pande, and J. Andrew McCammon, Accelerated Molecular Dynamics Simulations with the AMOEBA Polarizable Force Field on Graphics Processing Units, Journal of Chemical Theory and Computation *2013 9 (11), 4684-4691. Copyright 2013 American Chemical Society.*

BY KATHARINE MILLER AND STEPHANIE SIDES

A Boost for MD Sampling

In 2011, Simbios researchers reported achieving greater speed and accuracy in molecular dynamics simulations by tying the polarizable force fields from AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) to OpenMM, the Simbios toolkit that leverages the power of GPUs to accelerate simulations. Despite those gains, additional speedups are desired to generate simulations of these polarized molecules over ever longer timescales—reaching micro- and milliseconds.

A collaboration between **Steffan Lindert, PhD**, a postdoc from the University of California, San Diego, his advisor, **J. Andrew McCammon**, and Simbios researchers, has made new inroads by implementing accelerated molecular dynamics (aMD) with the powerful AMOEBA/OpenMM combo. aMD energetically raises regions of the potential energy landscape that fall below a certain cutoff. This lowers barriers between energy wells, allowing more frequent transitions between low energy states and resulting in enhanced sampling of the conformational space. Until Lindert came along, aMD had been implemented in classical (nonpolarizable) simulations in AMBER and NAMD, but not in OpenMM. "OpenMM was the ticket to GPUs and better performance," Lindert says.

Lindert spent a month last year as a OpenMM visiting scholar working with **Peter Eastman, PhD**, in the lab of **Vijay Pande, PhD**, at Stanford University. And the collaboration paid off: Lindert and the Simbios team found that the synergies between the aMD method and the AMOEBA force field conserve AMOEBA's accuracy while improving sampling efficiency by two to three orders of magnitude. These results, which were published in the *Journal of Chemical Theory and Computation* in October 2013, suggest that the aMD sampling method as implemented in OpenMM should effectively support studies that require more extensive (micro- to milli-second) biomolecular sampling. □

DETAILS

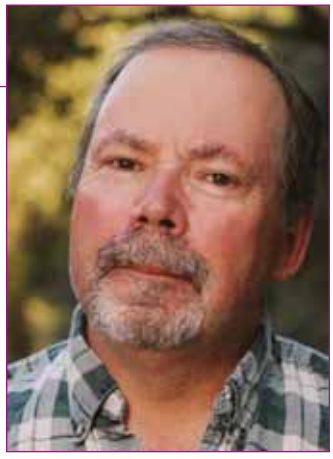
The AMOEBA-aMD implementation takes full advantage of GPU computing and is publicly available in OpenMM at http://wiki.simtk.org/openmm/VirtualRepository.

To learn more about the 2014 OpenMM Visiting Scholars program, visit http://simbios.stanford.edu/OpenMMVisitingScholar.

CommunityNews

BY KATHARINE MILLER

Philip Bourne Named as NIH Associate Director for Data Science



In December 2013, after a nearly yearlong search, NIH Director **Francis S. Collins, MD, PhD**, appointed **Philip Bourne, PhD**, to fill a permanent position as Associate Director for Data Science (ADDS) starting in January 2014.

"By devoting a high-level position to leading and coordinating strategies related to computation and informatics in biomedicine, the NIH signals a recognition that the era of Big Data in biomedicine has arrived," says **Russ B. Altman, MD, PhD**, professor of bioengineering, genetics and medicine at Stanford University and principal investigator of Simbios, the National Center for Physics-Based Simulation of Biological Structures. "This constitutes yet another step towards having biomedical research and medical practice become more quantitative, precise, and reproducible."

Bourne is also a great pick for the job, Altman says, given his wide range of relevant experience, including helping to lead the creation of a major biological database (the Protein Data Bank); using data mining techniques for discovery; starting an open-source journal; creating novel ways for delivering biomedical content using the Internet, and many other innovations. "He also has just the right temperament for navigating the complex structures at NIH," Altman adds.

Bourne's deep technical understanding of the issues in biomedical computing will also serve him well, Altman says. "This is a technical discipline and the leadership must have technical credentials. We are in good hands."

For more information on the Bourne appointment, see the NIH press release at http://www.nih.gov/news/health/dec2013/od-09.htm □

BIOLOGY:
A Game for a Crowd

By Sarah C.P. Williams

The rules of Phylo are simple: drag colored blocks across rows on the computer screen until similar colors line up. Within minutes of launching the game, any average person can learn how to play and begin developing strategies to beat the current best score, which is posted prominently at the top of the page.

It can be addictive even for players who don't know that the colored blocks actually represent gene sequences submitted by scientists trying to solve real-world biological puzzles.

"I wanted a game that an average person could play when they had a couple minutes—like Tetris—but that would be useful for bioinformatics," says **Jerome Waldispuhl, PhD**, an assistant professor of computer science at McGill University who spearheaded the development of the game with a colleague.

Phylo is helping Waldispuhl tackle the challenge of what—in biological jargon irrelevant to the gamers—is called multiple sequence alignment. "When a geneticist gets a new sequence, the first thing they often want to do is compare it to sequences from other species or individuals," he explains. Comparing sequences means lining them up, finding bits of the genetic code that match between the samples. Bioinformaticians typically rely on computer programs to parse the data and come up with such an alignment. But the solution provided by the computer, based solely on statistics, isn't always the best alignment, Waldispuhl says. "Multiple sequence alignment is a problem that is very difficult in computer science," he says. "But it's also one of the most used techniques in genomics studies today."

To improve on what the computers do, geneticists usually sift through the data manually, looking for ways to rearrange chunks of nucleotides to match up with sequences in other samples. Waldispuhl realized that when geneticists worked through the problem in this manual way, their knowledge of genetics wasn't itself vital; they viewed the chunks of genes as they might view colored blocks in a puzzle. That realization led to Phylo.

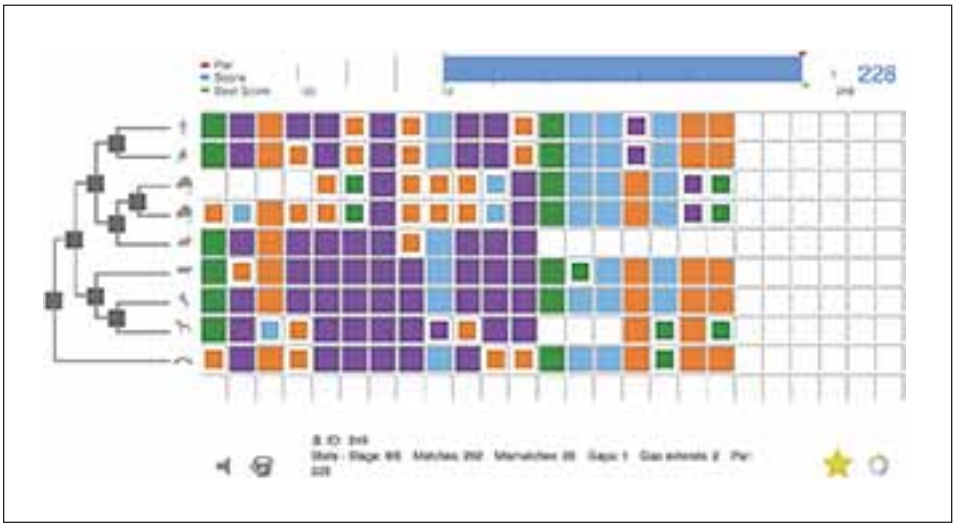
Phylo isn't the first, or the most-played, game that aims to solve scientific puzzles that have stumped researchers and computers. But it's part of a growing trend to drive biology forward by initiating games and competitions—among scientists and non-scientists alike. Foundations are offering cash prizes to those who come up with the best solution to a scientific quandary; institutes are posing broad research questions to people across disciplines to encourage out-of-the-box thinking; and computer scientists are teaming up with life scientists to turn biological enigmas into games for the average public.

> "I wanted a game that an average person could play when they had a couple minutes—like Tetris—but that would be useful for bioinformatics," says Waldispuhl.

Waldispuhl—like others involved with such initiatives—is careful to make it clear that the games and competitions aren't replacing the important work being done by skilled scientists; they're supplementing this work.

"I wouldn't say that humans do better than computers at every part of this task," Waldispuhl says of multiple sequence alignment. "What we tried to do with Phylo is find a better synergy between what humans can do and what computers can do."

More than 300,000 people have played Phylo since it launched in 2010, and Waldispuhl's team reported in a 2013 *Genome Biology* paper that up to 50 percent of the time, a casual gamer can match the performance of expert players; and up to 40 percent of the time, Phylo players can improve on the solution found by a computer program. Now, the scientists have expanded the game so that researchers working on any genetic problem—from high blood pressure to can-



In the game of Phylo, players try to align colored blocks representing genetic sequence information for different species. Image captured from phylo.cs.mcgill.ca

cer—can submit their sequences to be aligned by gamers. The players' alignments don't cure disease, but they provide better starting points for the researchers who do.

Shedding Scientific Baggage

Dan MacLean, PhD, a bioinformatician at The Sainsbury Laboratory in Norwich, the United Kingdom, had a similar goal in mind when he developed the game Fraxinus. Rather than lining up blocks of colors, Fraxinus players line up green, orange, yellow, and red tinted ash leaves that represent nucleotides. The symbolism is purposeful: the genetic sequences in this game are from ash trees and the fungus (*Chalara fraxinea*) that threatens up to 95 percent of Britain's ash trees. MacLean and his colleagues are racing against the clock to understand this fungal disease—ash dieback—and save the country's trees. They want to know what parts of the fungus genome make it so effective at causing disease, as well as whether particular genetic sequences in the ash trees make them more or less susceptible to infection.

Fraxinus, like Phylo, involves matching sequences of colored shapes to a given pattern. The fact that the pattern represents the genetic sequence of an ash tree is irrelevant to the gamer. Image captured from https://apps.facebook.com/fraxinusgame/

The challenge in developing a game to help compare genetic sequences of different ash trees and different fungi, MacLean says, was to understand what basic rules scientists followed when they normally analyzed the data and how to convey those rules in the game.

"It was really a matter of working out what was the scientific baggage I was carrying around, what were the assumptions I was making," he says. Working with a team of non-biologists helped him pare the problems down to a simple game. In the first 10 weeks Fraxinus was online, more than 10,000 sequences were analyzed.

"A computer program makes certain assumptions and there are certain limits on the number of permutations it can try," MacLean says. "These sequences we put into the game are not the low-hanging fruit; they're not the easy ones to solve. They're the ones that have multiple gaps and multiple possible arrangements. For these sorts of things, humans can do it better than computers."

> "These sequences we put into the game are not the low-hanging fruit; they're not the easy ones to solve. They're the ones that have multiple gaps and multiple possible arrangements. For these sorts of things, humans can do it better than computers," MacLean says.

MacLean hopes that by observing how people successfully solve puzzles within the confines of the Fraxinus game, he can learn better ways to teach computer programs to align the sequences. The very lack of scientific baggage that the players have, he says, may help players come up with new approaches to the problem. The scientists are still reviewing the data analyzed by the players, but hope that genetic mutations linked to ash dieback susceptibility will be revealed.

Other online games are also advancing science with assistance from creative non-scientists. Since 2008, players of Foldit have helped determine how proteins fold; while Eyewire users trace the route of a neuron's axons through MRI slices to help build the connectome.

Open Competitions

But games aren't the only way scientists are crowdsourcing computational work to unstick the way scientists think. Universities and other entities have been launching idea challenges, often with a cash prize attached.

For example, since 2010, researchers involved in Harvard Catalyst—a cross-disciplinary effort to drive biomedical research forward—have hosted several open challenges. In February 2013, **Eva Guinan, MD**, a radiation oncologist and director of the Harvard Catalyst Linkages Program, and colleagues at Harvard, announced the re-

sults of a complex immuno-genomics challenge. The competition, which carried weekly $500 prizes, sought a program that could more quickly analyze vast amounts of sequence data for the genes that make antibodies and T-cell receptors (TCRs). It's a tough problem because unlike other genes, those for antibodies and TCRs are built up combinatorially in each cell—i.e., they differ from cell to cell—making it difficult to trace the genetic origin of particular antibodies or TCRs. To issue the challenge, Guinan's team uploaded genetic data onto the website TopCoder and rephrased their problem in generic, non-biological terminology. They essentially created an information-theory and string-processing task that any computational expert could tackle.

The results astounded Guinan: The coders who entered the contest didn't converge on a single best method to solve the antibody challenge. Instead, out of more than 100 submissions, 16 different new approaches worked better than the standard algorithm. And they didn't just give better solutions; some were nearly a thousand times faster too.

"Our small community had missed not just one opportunity to improve the way we were doing things, but many different opportunities," says Guinan. Now, scientists can integrate the new methods into the way they study antibodies.

"When you go to that many people, you have some who are good at writing algorithms, some who are good at thinking about statistics, some who are good at string theory," says Guinan. "You get this amazing diversity of repertoire in the solver population. How could I possibly hire that many different experts here?" Guinan once again emphasizes that such competitions don't detract from the work that scientists themselves do. In fact, she says, it takes skilled scientists to design and implement an effective competition.

"A solution is only going to be good if the question posed is good," she says. "And that's where the scientist comes in. This isn't just about throwing a bunch of data at someone and saying 'Call me when you have the answer.'" The Harvard Catalyst group now has more competitions in the works. Up next: challenges to address the genetics of HIV and the best way to analyze colonoscopy data.

"A solution is only going to be good if the question posed is good," Guinan says. "And that's where the scientist comes in. This isn't just about throwing a bunch of data at someone and saying 'Call me when you have the answer.'"

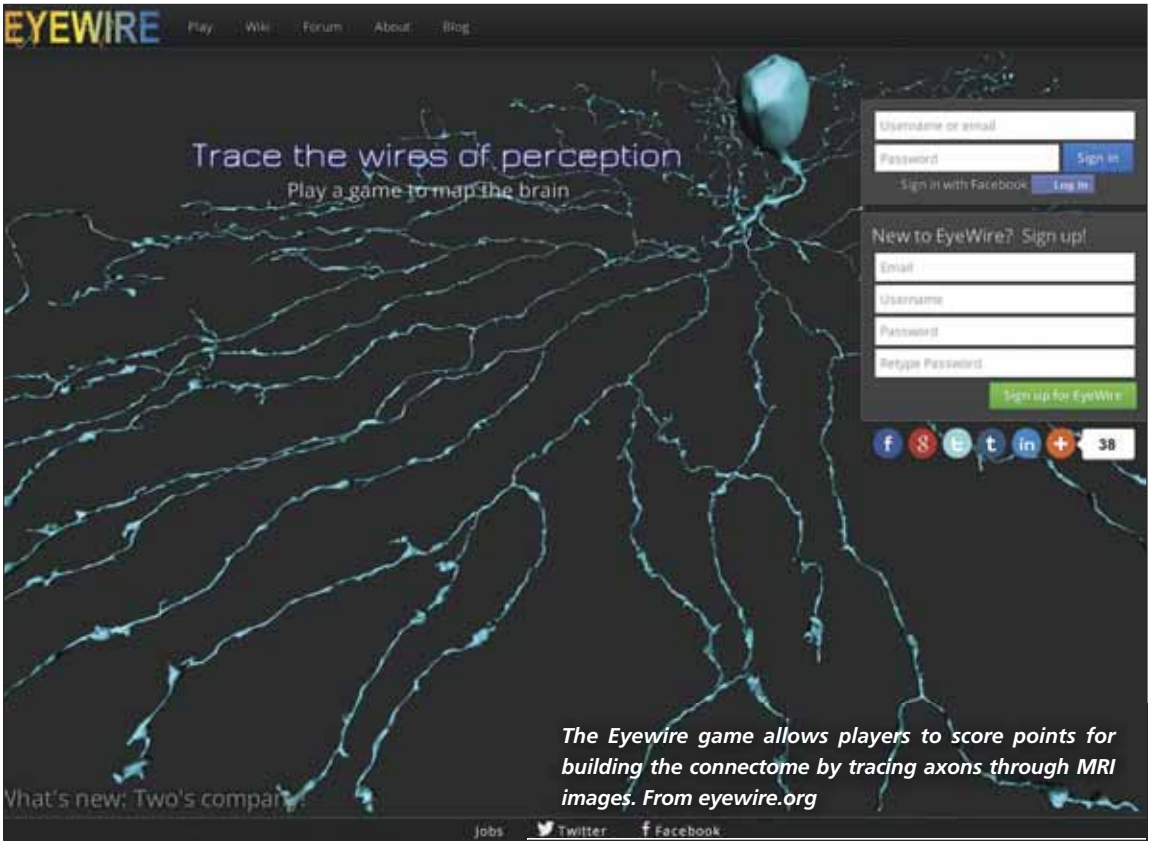
Sweetening the Pot

Some open challenges offer even bigger prizes. In 2011, for example, the Pistoia Alliance announced the Pistoia Sequence Squeeze Competition to develop new ways of compressing genetic data, with a $15,000 prize for the best solution. **Richard Holland**, the chief business officer of Eagle Genomics, teamed up with Pistoia to run the competition. He says that updating results constantly—with a leaderboard—was technically challenging but helped spur faster improvements.

"It encourages people to compete in order to outdo each other," he says. Leaders constantly leap-frogged each other as they improved their techniques. Holland also learned that having strictly defined entry and judging criteria ensures that a contest runs smoothly.

And the prize money didn't hurt. For scientists and non-scientists alike, the drive to win a game or competition—whether for money or pride—combined with the natural human instinct to solve puzzles can be a powerful motivator in driving research forward.

"The wrong message here would be to say that humans are better than computers or that crowd-sourcing is better than experts," says Waldispuhl. "It's more about trying to see where we can improve different steps of the research process by doing things a little bit differently." □



The Eyewire game allows players to score points for building the connectome by tracing axons through MRI images. From eyewire.org

i2b2 GOES VIRAL:
Open-Source Platform Enables Clinical Research

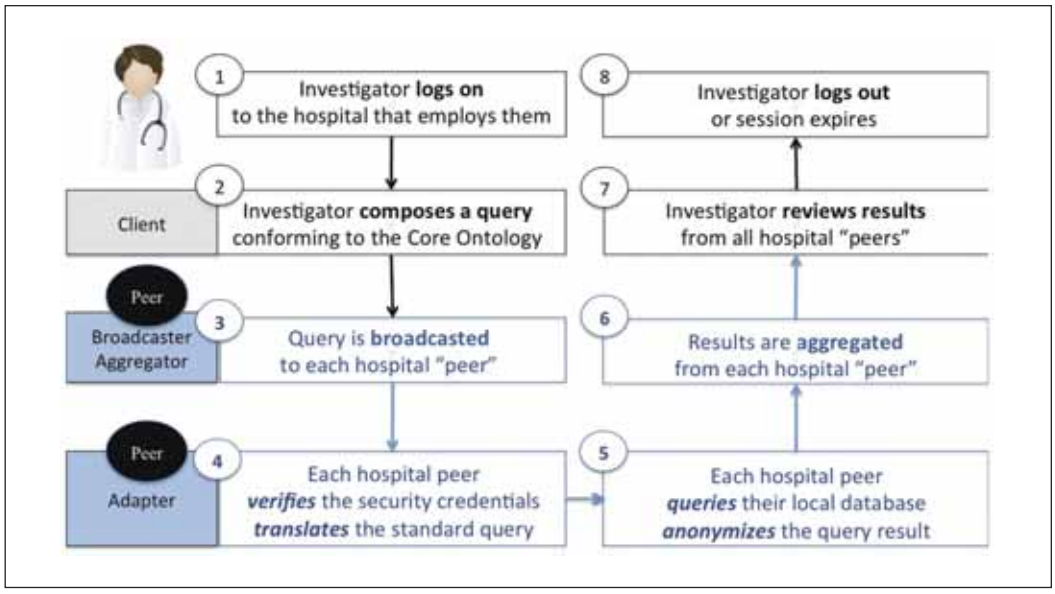
By Kristin Sainani

In 2011, the FDA issued a black box warning on Celexa, one of the most widely prescribed antidepressants in the US. At higher doses, the drug had been linked to potentially dangerous changes in the electrical activity of the heart (a so-called prolonged QT interval). As a result, physicians began switching patients to a similar drug called Lexapro. "It was absurd because Lexapro and Celexa are almost exactly the same chemically," says **Shawn Murphy, MD, PhD**, associate professor of neurology at Massachusetts General Hospital. "But because the FDA did not conduct a study of patients taking Lexapro, it didn't have a black box warning."

Rather than wait a few years for the FDA to gather new data about Lexapro, Murphy and his colleagues decided to look at data that already existed in electronic medical records. And to do that, they turned to the i2b2 open-source software suite. Developed by i2b2 (Informatics for Integrating Biology and the Bedside), a National Center for Biomedical Computing (of which Murphy is a part), the i2b2 platform enables researchers to use existing clinical data for discovery research.

To look at Lexapro's side effects, the team mined electronic medical records from more than 38,000 patients with both antidepressant prescribing data and heart monitoring data (EKGs) available. "And we could actually show—just based on these previously acquired EKGs—that Lexapro was not only causing a prolonged QT interval, it was even a little bit worse than Celexa," Murphy says. The study was published in *BMJ* in 2013 and the results are being reviewed by the FDA. "That was an adverse event that we could detect just by going back and mining the electronic data that had been collected in the normal course of clinical care on our patients," Murphy says.

It's just one of a number of recent success stories from i2b2's creators, who have used the platform for everything from identifying adverse drug events to performing genome-wide association studies to discovering novel subclasses of diseases—all at a fraction of the time and cost of conventional studies. But the impact of i2b2 has spread far beyond Harvard's hospitals; it has been widely adopted elsewhere. "We know of 120 hospitals that have implemented it. But it could be double that, since people don't have to report to us," Murphy says.



Data Without Borders: The i2b2 SHRINE (Shared Health Research Information Network) network enables hospitals to pool their data without compromising patient or institution privacy. Reprinted from McMurry, AJ,* et al., *SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies* PLoS One *DOI: 10.1371/journal.pone.0055811 (2013).

Institutions are moving to i2b2 in large numbers because of its proven track record; large user group; and open-source license.

"There's not a general competitor for the i2b2 platform only because the terms of i2b2 are so liberal," Murphy explains. "The most rewarding and unique thing about i2b2 is that it really is just the starting point for people who want to develop their own tools at their own hospital." Cancer centers, for example, might use their own data to develop their own survival plots—graphs showing the long-term benefit of various treatments or the prognosis for people with different cancer types—while children's hospitals might want to build growth curves. "They're totally enabled to develop it in ways they see fit for their use cases," Murphy says.

To capture a snapshot of i2b2's impact, *Biomedical Computation Review* talked to informatics leaders at two hospitals that have been active in the i2b2 community: the

"There's not a general competitor for the i2b2 platform only because the terms of i2b2 are so liberal," Murphy says.

University of Kansas (KU) Medical Center and the Cincinnati Children's Hospital.

Kansas: Extending i2b2 and Connecting to Other Institutions

In 2010, when **Russ Waitman, PhD**, was recruited to be KU's director of medical informatics, he made the critical decision to build the university's clinical research informatics infrastructure on i2b2. Given a limited budget, he says, "I couldn't afford to sit here and be yet one more informatics shop that's going to reinvent the wheel." Instead, he decided to "take the wheel from Harvard and build a car."

i2b2 allows investigators to identify cohorts of patients for research studies in a self-serve manner. Researchers can do limited analyses on the de-identified data and if the results are promising can then request approval from their institutional review board to obtain the full dataset. In 2013, the KU system served 142 users with 4,751 queries; and fulfilled 53 further requests for full datasets. Waitman's team has also expanded on what i2b2 can do. For example, they can now integrate hospital data with data from cancer registries; and statistical analysis plugins they created allow users to analyze datasets on the server, without having to download them to their computers (thus protecting the data from potential release or loss of privacy protections).

Waitman's team is leading an effort to share data across multiple hospitals using the i2b2 platform. They recently received a Clinical Data Research Network contract (http://www.pcornet.org/clinical-data-research-networks/) to link ten health systems that are already using i2b2 in Kansas, Iowa, Missouri, Wisconsin, Minnesota, Texas, and Nebraska (the Greater Plains Collaborative). Waitman is excited about this, he says, because researchers will be able to perform queries across records for the 6 million patients at these institutions.

Using the i2b2 platform, electronic medical records can also be linked to biological samples collected during routine hospital and clinic visits. Initially, Waitman's team hopes to work with surgical pathologists to determine whether samples held by their hospitals could be useful for research. Waitman's team is also looking at ways to use the data stored in i2b2 to drive hospital quality improvement. For example, data from electronic medical records (EMR) could reveal how well different doctors are meeting goals for diabetic control. "Because that's what I think is going to make it really sustainable," Waitman says. "You have your core clinical research covered with your EMR data and then you want to blend downstream to the biological and then you want to blend upstream to get the health systems benefiting from it."

Cincinnati: Boosting i2b2 User-friendliness and Building Disease Registries

"We've had our hands dirty with i2b2 for quite a while now," says **Keith Marsolo, PhD**, associate professor of pediatrics. When he was recruited to Cincinnati Children's in 2007 to build a research data warehouse, he surveyed available platforms. Some institutions like Stanford had built their own custom systems; and IBM had a commercial product. But i2b2 was the go-to choice for open source. "It appealed to us because it was open source [which meant] we could tinker with it and build new things on top of it."

"We've had our hands dirty with i2b2 for quite a while now," says Marsolo .... "It appealed to us because it was open source [which meant] we could tinker with it and build new things on top of it."

Marsolo's team has built a number of plugins for i2b2 that they have also made available to the i2b2 community. "Marsolo has taken it in a lot of interesting directions that we hadn't even conceived of," Murphy says. His team built the first web browser for i2b2; an application for viewing clinical data in a web-based form (similar to a chart review); and a forms module to allow direct data entry into i2b2. They are also linking clinical data to biological specimens and genetic data.

Marsolo's team is also working on integrating natural-language processing into their i2b2 pipeline using the open-source software cTAKES (clinical Text Analysis and Knowledge Extraction System). Much of the relevant data in electronic medical records is locked away in narrative notes. But open-source text-mining tools such as cTAKES can extract critical information from free text.

And Cincinnati investigators are already using the system to make novel discoveries. In a December 2013 paper in the *Journal of Pediatric Gastroenterology and Nutrition*, Marsolo and colleagues showed, for the first time, that two rare childhood diseases—eosinophilic gastrointestinal disorders (EGID) and pediatric PTEN hamartoma tumor syndromes (PHTS)—are strongly associated. By querying their i2b2 data warehouse and a local eosinophilic database, they were able to search through more than one million patients to find eight with confirmed PHTS, five of whom also had EGID.

Marsolo's team has also focused on creating multi-center disease registries using i2b2's data sharing software SHRINE (Shared Health Research Information Network). Chronic childhood diseases are often hard to study due to the small number of patients at any single institution. But SHRINE allows institutions to share data while protecting patient privacy. Marsolo's team helped create the 60-site CARRAnet (Childhood Arthritis and Rheumatological Research Alliance); with more than 5,000 patients, this registry is the largest available for pediatric patients with rheumatic disease.

Cincinnati Children's is also one of two children's hospitals involved in the eMERGE (electronic Medical Records and Genomics) network, which links electronic medical record data with genetic samples (and also uses i2b2). Using data from this network, Marsolo and colleagues performed a genome-wide association study looking for genetic variants associated with obesity in children. Their results confirmed previous results from adult studies and also identified novel variants in kids. The results were published in *Frontiers in Genetics* in December 2013.

The future of i2b2 may depend on how electronic medical record companies evolve their tools going forward, Marsolo says. For example, EPIC, a medical software company, has recently come out with a tool for identifying cohorts of non-anonymized patients. Even if a de-identified version of that tool "might change the value proposition and primary use case for i2b2," Marsolo says, i2b2 will likely endure because it is customizable and fosters data sharing across multiple institutions. □

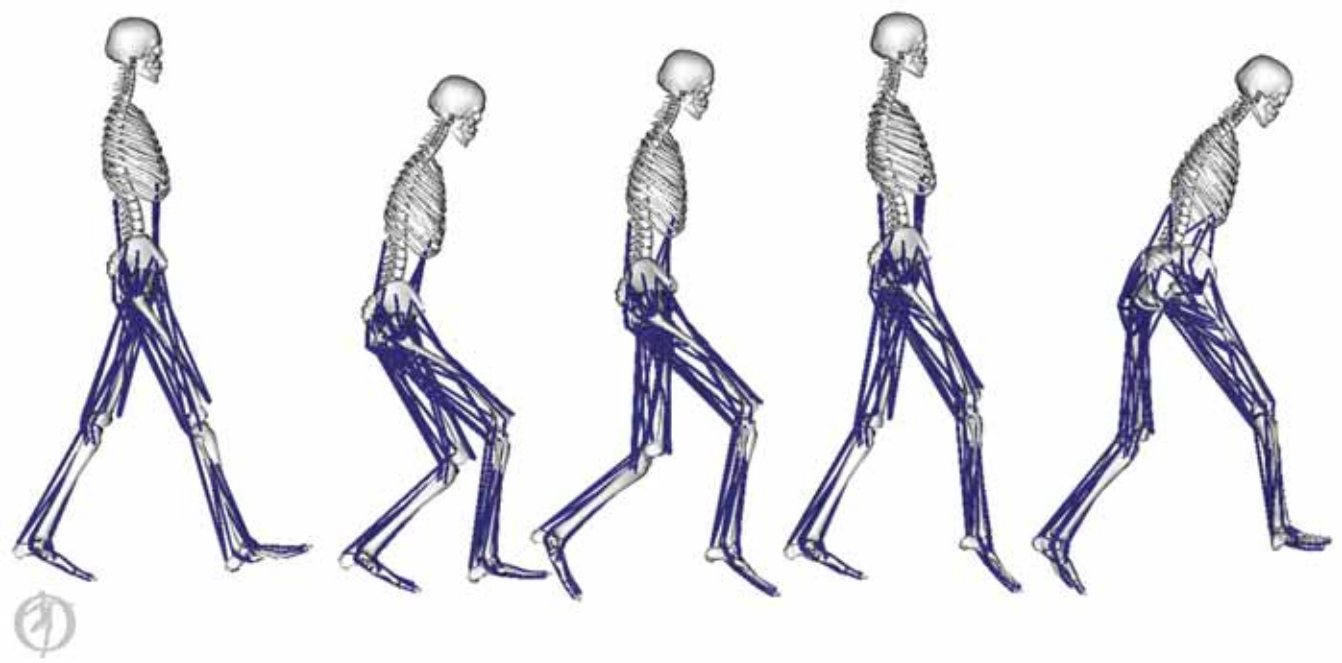
COMPUTATION IN THE SURGICAL SUITE:
Modeling Crouch Gait for Orthopedic Decision Making

By Alexander Gelfand

They say that what you don't know can't hurt you. But what your pediatric surgeon doesn't know certainly can hurt your child. An orthopedist who recommends that a child with cerebral palsy undergo hamstring surgery to correct a gait problem, for example, often can't be sure if the procedure will make things better, or worse.

Here, we tell the tale of how one team of researchers and clinicians is using computers to help correct crouch gait, a problem that afflicts many patients with cerebral palsy.

Michael Schwartz, PhD, and **Tom Novachek, MD**, have been working hand in surgical glove for many years to integrate computational methods into the surgical decision-making process. Both men work at the James R. Gage Center for Gait and Motion Analysis at Gillette Children's Specialty Healthcare in St. Paul, Minnesota—Novachek as the Center's medical director, and Schwartz as its director of bioengineering research. (Both also hold appointments in the department of orthopedic surgery at the University of Minnesota.)



In these images generated by the OpenSim software, an unimpaired individual stands upright (far left figure), while four different individuals exhibit different types of crouch gait, each with different underlying causes. Using OpenSim, surgeons can better understand which patients with crouch gait will benefit from surgery to lengthen the hamstrings. Courtesy of Jennifer Hicks, executive director for OpenSim, Stanford University.

Physicians and patients confront scenarios like this every day, with doctors making their best guesses about what might work, and those in their care trusting that everything will go according to plan. To boost their chances of being right, surgeons in various orthopedic specialties are turning to computational modeling and simulation to better predict patient outcomes and to improve accuracy and safety in the operating room. Some, for example, use modeling software to more accurately fit knee and hip implants to bone, while others go so far as to rely on computer-guided robots to do some of the cutting and sanding for them.

The Center uses the same motion-capture technology employed in Hollywood films to diagnose and plan treatments for people with walking and movement disor-

ders. Approximately three quarters of the patients who enter its doors have cerebral palsy. Of those, some 90 percent are children, many of whom walk and run with their knees perpetually flexed, a painful and potentially debilitating condition known as crouch gait.

Because crouch gait is often attributed to excessively tight hamstring muscles, surgeons have typically tried to correct the disorder by lengthening them, a process that involves elongating the tendons that attach the hamstrings to the bones. But factors aside from hamstring length, such as neurological defects and bone deformities, may also contribute to crouch gait, and doctors have historically lacked good tools for predicting who will respond well to the procedure. As a result, while some young patients see improvement after surgery, others get worse; and even those who make headway may require further corrective surgery later on in life.

Since the mid-1990s, motion-capture hardware and 3-D simulation software have improved the situation somewhat by giving doctors a better understanding of how bone deformities and excessive tightness or looseness in other muscles can also lead to crouch gait. At Gillette, for example, every patient is outfitted with reflective markers and videotaped as he or she moves across the motion-capture lab. Commercial software uses the trajectories of the markers and the angles of the patient's joints to generate an animated 3-D simulation that can be used to analyze the individual's joint kinematics. Doctors can then use that information along with other patient-specific data to figure out what's driving a particular problem. As sophisticated as it is, however, the system still has limitations; so Schwartz and Novachek have added another tool to their technological arsenal: OpenSim, the open-source software package for modeling and simulating movement developed through Simbios and the NIH Center for Simulation in Rehabilitation Research (NCSRR) at Stanford University.

In a series of articles, a group of researchers led by **Scott Delp, PhD**, professor of bioengineering, mechanical engineering, and orthopedic surgery at Stanford, used musculoskeletal models developed in OpenSim to demonstrate that the lengths and velocities of a patient's hamstrings could be used to predict whether they would benefit from hamstring lengthening. (The velocity of a muscle describes the rate at which its length changes over time.) Schwartz, who contributed to some of those articles, explains that if the hamstrings are short enough and slow enough, then lengthening them can alleviate crouch gait. If not, lengthening may do nothing—or it may make matters worse.

Yet standard motion-capture gait analysis only offers an indirect measurement of muscle lengths and velocities; and basing surgical decisions on that kind of inference, says Novachek, involves "a little bit of a leap of faith." Fortunately, OpenSim can use the data gathered during motion capture to compute muscle lengths and velocities directly. So now, marker trajectories and joint angles that are recorded for a child with crouch gait are fed into OpenSim, which spits out its best estimate of the lengths and velocities of their hamstrings. The software is even capable of minimizing the errors that can arise from differences between the observed marker positions (which can be thrown off by inaccurate marker placement or soft tissue motion) and what it predicts those marker positions should be based on the recorded joint angles.

Powerful as it is, OpenSim can't yet tell Novachek and his fellow surgeons exactly how much a patient's hamstrings ought to be lengthened. Nonetheless, simply identifying those patients who are genuinely good candidates for the procedure has already bumped up surgical success rates, and Novacheck has also seen reductions in under- and over-corrections, the latter of which is especially difficult to treat. More generally, Novacheck says that the confidence afforded by gait analysis has made the Gillette surgeons more aggressive about the number of gait-related corrections they're willing to tackle during a single session—performing anywhere from five to fifteen at once, for example, rather than just one or two—thereby saving their patients from having to undergo multiple follow-up surgeries.

In the near term, Novacheck looks forward to seeing similar data on the lengths and velocities of other muscles implicated in crouch gait. In the long term, he hopes to one day bring motion-capture equipment right into the operating room in order to precisely document just how much hamstring lengthening is actually required to normalize a person's gait—information that should enable OpenSim to predict how many millimeters or centimeters need to be trimmed from a given patient's hamstrings before they go under the knife. □

Marker trajectories and joint angles that are recorded for a child with crouch gait are fed into OpenSim, which spits out its best estimate of the lengths and velocities of their hamstrings.

Novacheck says that the confidence afforded by gait analysis has made the Gillette surgeons more aggressive about the number of gait-related corrections they're willing to tackle during a single session.

COMPUTATION IN THE SURGICAL SUITE:
Navigating the Brain

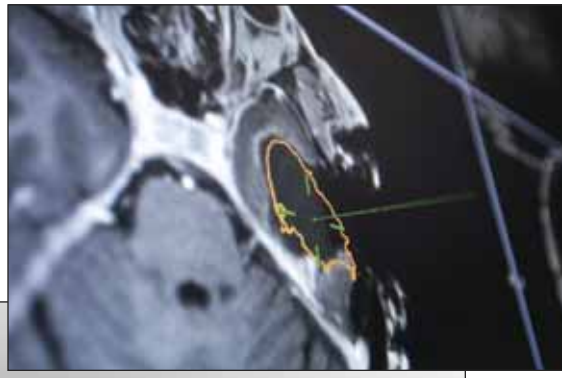
By Alexander Gelfand

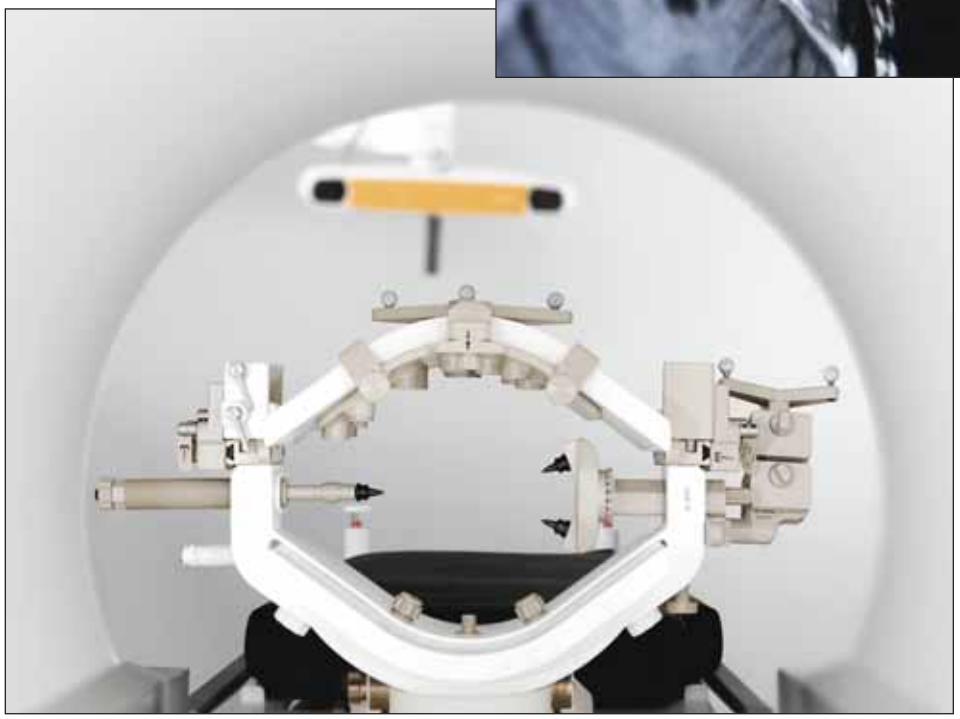
No surgical specialty has embraced computer technology more rapidly, or benefited more from it, than brain surgery. And with good reason: The brain does not readily yield its internal structure and function to the unaided eye, and a scalpel aimed a hair's breadth off course can mean the difference between miraculous recovery and personal catastrophe. As a result, neurosurgeons are keenly interested in anything that will help them operate in a minimally invasive manner and avoid collateral damage. Or as **Sujit Prabhu, MD**, professor of neurosurgery at the University of Texas M.D. Anderson Cancer Center in Houston, Texas, puts it: "My job is to prevent complications."

Navigational Software

Prabhu, who holds a joint appointment at Baylor College of Medicine, has been receiving assistance on that front from both a German medical technology firm and an American software company. For its part, Munich-based Brainlab is a major provider of surgical navigation systems that allow doctors to more effectively see what they're doing in the operating room. M.D. Anderson, for instance, uses an integrated system called Brainsuite® developed by Brainlab. It is comprised of a stereoscopic infrared camera yoked to dual high-definition monitors and some very sophisticated software. The camera detects infrared reflective markers that are attached to the patient and on Prabhu's instruments, while the software triangulates the relative spatial positions of both patient and tools and displays them on the screen.

Better yet, the software merges that tracking information with medical imaging data derived from the patient's preoperative computed tomography (CT), magnetic resonance imaging (MRI), and diffusion tensor imaging (DTI) scans, all of which can be overlaid on top of one another. Linear registration algorithms align the scans by matching up anatomical features—the tip of the nose, the tragus of an ear—or additional markers stuck to the patient's scalp, and applying various spatial transformations (rotation, reflection, scaling) to fit the images to one another. Once those images have been registered to the patient's anatomy—typically, by pointing a tracked tool at a few features on the patient's skin and identifying the same fea-





Brainlab Brainsuite® iMRI system is used for navigated neurosurgery at the M.D. Anderson Cancer Center in Houston. Photographs owned by Brainlab AG, all rights reserved.

Published by ***Simbios****, the NIH National Center*

tures in the scans—Prabhu can touch any spot on the patient's brain, and the software will display crosshairs onscreen directly above whatever anatomical structure (lobe or ventricle, gyrus or fiber tract) he's pointing at. Or at least, it would if those structures didn't move.

Prabhu can touch any spot on the patient's brain, and the software will display crosshairs onscreen directly above whatever anatomical structure (lobe or ventricle, gyrus or fiber tract) he's pointing at. Or at least, it would if those structures didn't move.

The Brain as a Moving Target

Unfortunately, due to a phenomenon known as brain shift, brain structures rarely remain in the same positions they occupied when the preoperative scans were taken. Not only can the brain swell and deflate like a crenulated grey balloon during surgery, says **Louis Collins, PhD**, professor of neurology, neurosurgery, and biomedical engineering at McGill University, it also has the consistency of medium-density tofu, making it prone to movement when poked or prodded. All of this means that while navigation systems can theoretically offer accuracy to within two millimeters or less, in reality, things are less clear-cut.

Surgeons have begun to compensate by using intraoperative scans. M.D. Anderson, for example, has built an operating suite with an integrated MRI scanner that can update a patient's images with more accurate data that takes into account the brain's movements. Collins, who leads the Image Processing Lab in the McConnell Brain Imaging Center at the Montreal Neurological Institute, has been using ultrasound technology for similar purposes. Aligning preoperative and intraoperative scans requires the use of more complex nonlinear registration algorithms that can warp one image to fit another, but such techniques are already finding their way into prime time: Collins has incorporated them into his prototype system in Montreal, and **Uli Mezger**, clinical research manager at Brainlab, says that the company will soon bundle "morphing" or "elastic fusion" algorithms into its commercial platforms.

Aligning preoperative and intraoperative scans requires the use of more complex nonlinear registration algorithms that can warp one image to fit another, but such techniques are already finding their way into prime time.

Steering Around Brain Functions

Knowing where the anatomical structures in a patient's brain are located is only half the battle. The other half involves determining their function, and predicting how slicing through them might affect the patient lying on the table before you. That's especially true for a surgeon like Prabhu, who specializes in removing the malignant tumors known as gliomas from those areas of the brain that control speech, movement, and the senses.

Direct electrical stimulation is the most accurate way of determining function, but surgically inserting electrodes in a patient's brain is both time-consuming and invasive. Functional MRI (fMRI) and positron emission tomography (PET) scans are noninvasive and relatively quick; but even when those technologies are available (and not every medical facility can offer them), they don't necessarily provide the precision or spatial resolution that a surgeon needs. Nor can the information they supply always be closely matched to structural images like CT and standard MRI scans, in part because the tumors themselves can distort the brain's anatomy and interfere with the functional imaging process. "We struggle with concordance, especially when we're working in 3-D spaces," Prabhu says.

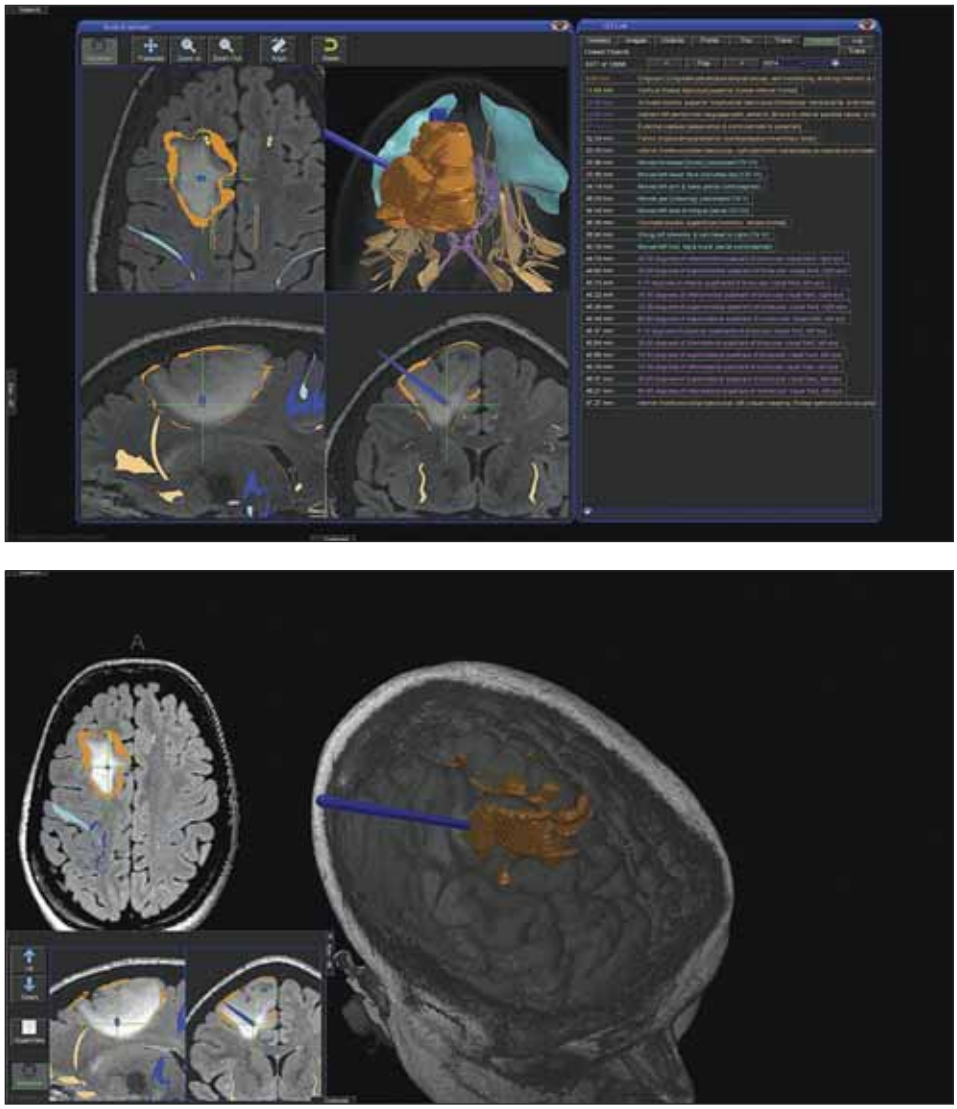
Recently, however, Prabhu has been working with a Houston-based startup called Anatom-e to overcome these hurdles. Over the past 10 years, the company has developed a 3-D model of the human brain in which every internal structure has been functionally annotated. According to **Mark Vabulas**, Anatom-e's CTO, the model—which is known variously as a brain atlas and a deformable anatomical template, or DAT—is spatially coherent. Consequently, any changes made to the shape of one sector ripple out to adjacent ones in a consistent and realistic fashion, allowing it to be reliably and accurately mapped to the scans of any given patient using a simple linear registration algorithm, then checked and, if necessary, tweaked by a human operator—a process that can be repeated with intraoperative scans to compensate for brain shift. Large tumors can distort local anatomy to such an extent that a user must manually register the atlas using visual landmarks, but the end result remains the same: a composite image of the patient's brain that ties structure to function at high resolution. **L. Anne Hayman, MD**, one of the company's founders, calls the DAT a "GPS for the brain," and it may indeed prove as useful to neurosurgeons as Google Maps is to the rest of us.

During preoperative planning to remove a glioma, Prabhu and his team can match the atlas to a patient's scans to outline the limits of the tumor, assess the degree to which it impinges on neighboring structures, and plan the best trajectory through the brain to reach it. In the OR, Prabhu can touch his tracked instruments to any part of the patient's brain and the atlas will supply its function, its distance to critical nearby regions, and a host of other useful information, all displayed in a multicolored 3-D image that can be rotated, expanded, and otherwise manipulated in a variety of ways. The technology is already producing results: Last year, Prabhu and Vabulas coauthored a paper in the journal *Neurosurgery* describing how a team of surgeons at M.D. Anderson used the DAT to help remove tu-

mors from three patients. And in an online demonstration for this author, Vabulas not only illustrated just how closely the atlas' identification of various functional areas in a patient's brain corresponded to the results of direct electrical stimulation, he also replayed the animated digital data from an actual glioma biopsy that employed a Brainlab navigation system, the DAT, and intraoperative MRI. Onscreen, one could see the slender, dark blue avatar of a biopsy needle sliding along a light blue trajectory in order to safely reach its target.

Vabulas is currently working on fully automating the process of mapping the DAT onto a patient's medical images with a smart, "adaptively deformable" registration algorithm that knows enough about the physical parameters of different brain regions (e.g., their tensile properties, their water content) to realistically warp the atlas so that it can match even the most distorted anatomy. The algorithm, which Vabulas would like to begin testing this year, could also alert surgeons to potential abnormalities by recognizing areas of the brain that simply cannot be fitted to the atlas. "By knowing what the DAT can do," Vab-

L. Anne Hayman, MD, one of the founders of Anatom-e, calls their deformable anatomical template (DAT) a "GPS for the brain," and it may indeed prove as useful to neurosurgeons as Google Maps is to the rest of us.

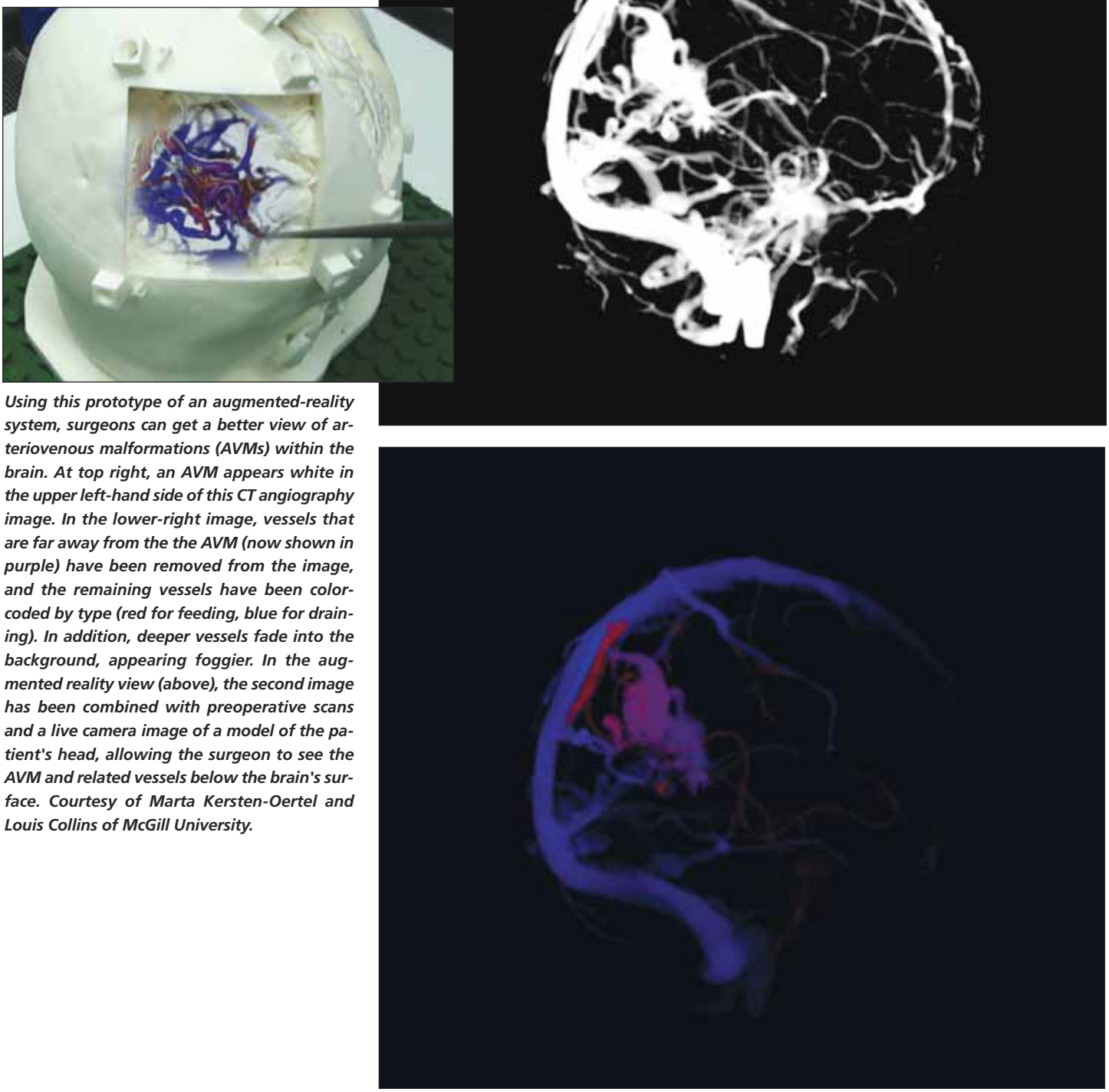


During intraoperative navigation while using the Anatom-e brain atlas, a surgeon might access images like these to help remove a tumor. The top image includes the three orthogonal views of the patient's brain (sagittal, axial, and coronal) with the tumor itself visible as a lighter-than-normal area surrounded by an orange outline or a solid orange mass. The surgeon's navigation probe is represented both as crosshairs and as a blue, 3-D wand; and the system provides a list of visible structures and their distances from the tip of the probe. The lower image, which includes a larger axial view and smaller sagittal and coronal ones, provides the surgeon with an idea of what to expect from a spatial perspective: In the large 3-D image on the right-hand side, one can see the relative positions of the blue probe, the tumor (in orange), and the surface of the patient's brain. Courtesy of Mark Vabulas and Anatom-e.

ulas says, "the algorithm can pinpoint spots that don't make sense."

Making Navigation Faster and Cheaper

Vabulas' DAT represents one way of using computational methods to give surgeons the information they need, when they need it—even if they don't have access to the most advanced imaging technologies. But there are other ways of leveraging registration algorithms, navigation systems, and multimodal imaging methods to give surgeons and their patients an edge in the OR.



Using this prototype of an augmented-reality system, surgeons can get a better view of arteriovenous malformations (AVMs) within the brain. At top right, an AVM appears white in the upper left-hand side of this CT angiography image. In the lower-right image, vessels that are far away from the the AVM (now shown in purple) have been removed from the image, and the remaining vessels have been color-coded by type (red for feeding, blue for draining). In addition, deeper vessels fade into the background, appearing foggier. In the augmented reality view (above), the second image has been combined with preoperative scans and a live camera image of a model of the patient's head, allowing the surgeon to see the AVM and related vessels below the brain's surface. Courtesy of Marta Kersten-Oertel and Louis Collins of McGill University.

For example, while intraoperative MRI can help compensate for brain shift, it is also, says Collins, extremely expensive: outfitting an operating suite with an intraoperative scanner and non-magnetic MR-compatible tools can cost upwards of several million dollars. It also takes time to prepare a patient for scanning and to execute and process the scans themselves, forcing surgeons to wait for as long as thirty or forty minutes before they can see precisely what's going on inside a person's head.

Seeking faster results at lower cost, Collins has turned instead to intraoperative ultrasound. An ultrasound scanner costs roughly $50,000; and when an ultrasound probe is placed on the cortex or inserted into the surgical cavity in a patient's brain, as many as 300 to 400 images can be acquired in less than a minute. Since the probe is tracked by an optical system that is accurate to less than one millimeter, Collins' navigation platform, which goes by the name IBIS (for Interactive Brain Imaging System), can determine the precise 3-D location of every pixel in the resulting images and create 3-D reconstructions of the patient's brain as it appears in surgery. The system can then employ linear and nonlinear registration techniques to align the patient's intraoperative ultrasound data with his or her preoperative scans (e.g., MRI and CT for anatomy, fMRI and PET for function), warping the latter to fit the former as necessary.

When Collins and his students first began developing IBIS a decade ago, it lacked the processing speed to perform those volumetric reconstructions and image registrations in the OR. Two years ago, they had the whole process down to 10 minutes; but that still wasn't fast enough. "They were quite nice and patient with us," Collins says of his surgical colleagues, "but they weren't really using the data." Now, however, thanks to a dedicated graphics processing unit, IBIS can process those requests so quickly—performing reconstructions in approximately one and a half seconds, and registrations in just one—that by the time a surgeon has completed an ultrasound scan and handed the probe back to a nurse, the results are already onscreen.

Collins believes this ultra-fast approach will change the way neurosurgeons work, allowing them to re-image patients far more often than they do now in order to safely remove as much tumorous tissue as possible. He is also investigating the possibility of using transcranial ultrasound to help surgeons insert exceedingly long, thin needles into the subthalamic nuclei of Parkinson's patients in preparation for deep-brain stimulation, a task he likens to sucking a specific seed from the center of a melon with a straw. (It's an apt analogy: the region of interest is roughly the size of a cantaloupe seed, and lies 8 to 10 centimeters inside the brain.)

The Right Information at the Right Time

As is often the case with new technologies, usability is an issue with computer-assisted surgery. On the one hand, researchers and developers can make things easier on clinicians (and drive down the risk of human error) by introducing more automation. On the other, they need to think carefully about the information they choose to present and how they present it, so as not to overwhelm or distract surgeons with an indiscriminate flood of data that's difficult to interpret or irrelevant to the task at hand. "The key," says Mezger, "is to display the right information at the right time."

"The key," says Brainlab's Mezger, "is to display the right information at the right time."

Those concerns underlie another one of Collins' projects: an augmented-reality system to help surgeons remove arteriovenous malformations (AVMs) from patients' brains. Left untreated, these abnormal tangles of blood vessels, which Collins likens to balls of knotted yarn, can cause headaches, epileptic seizures, and even strokes. Removing them can be tricky, however, since surgeons must first distinguish the vessels that feed blood to the AVMs from those that drain them, then sever them in the correct order, all while finding their way through opaque brain tissue. Preoperative CT, MRI, and angiography can map a patient's vessels; image-processing algorithms can sort feeders from drainers by tracing their connections back to major arteries and veins; and operating microscopes can provide close-ups of the vessels once they have been exposed. But how to deliver all of that visual information in a unified, comprehensible way?

Initially, Collins fed the combined imagery into clunky stereoscopic goggles, which no one much liked. Now, however, he and his team present it all on a single monitor. The system employs the same navigation platform, tracking system, and registration algorithms as the intraoperative ultrasound set-up to ensure that everything lines up properly on screen. In place of a tumor and its surrounding anatomy, however, the surgeon sees blood vessels that have been color coded both by type (red for feeding, blue for draining) and by depth, so that he can zero in on the ones that matter most—namely, those associated with the AVM itself, and those that are in the way. This "chromadepth" method of representing distance isn't perfect; while it provides a quantitative sense of the relative depths at which different vessels lie, they still look as if they are floating on top of the patient's brain. But Collins is already experimenting with ways of enhancing depth perception, such as making deeper vessels appear foggier, like distant figures in a landscape. The goal, he says, is to minimize the cognitive load imposed by the technology so that surgeons can concentrate on the job at hand.

The goal, Collins says, is to minimize the cognitive load imposed by the technology so that surgeons can concentrate on the job at hand.

If he succeeds, he'll be able to add his augmented-reality system to the growing list of technologies that are helping surgeons see their patients in a whole new light. □

Mobilizing BIG DATA to Understand Mobility

By Esther Landhuis

In a win-win for patients and researchers, big biomechanics data has arrived.

One rainy day in the late 1990s, a biomechanics doctoral student stood sopping at the street corner, hoping his research participant would arrive for her gait analysis. "It was the third time she stood me up. I was sick and tired of waiting for people to come to me," says **Reed Ferber, PhD**, recalling his graduate years at the University of Oregon. "I wanted to create something where people stream in all the time and we have access to their data at our fingertips."

That yearning eventually birthed the Running Injury Clinic, which Ferber now heads as an associate professor of kinesiology at the University of Calgary, Canada. Unlike his graduate research subjects who grudgingly "gave two hours of their time and didn't get much out of it," he says, hordes of runners have flooded into the clinic, eager to fork over several hundred dollars to learn how they got hurt and what they can do about it. It's a win for data-hungry researchers, too. Cameras record every twist and turn of the runners' ankles, knees, and hips as they walk or jog on treadmills. All that data flows straight into a vast 3-D motion-capture repository, which stores data from the Calgary clinic as well as some 30 gait analysis centers around the globe.

It's a sure sign that biomechanics labs are starting to tap into the big data revolution.

Just as Amazon uses big data to deduce what a consumer might crave, or genomics researchers use it to identify genes that cause a specific disease, so too can biomechanics researchers like Ferber benefit from big data—mining it to distinguish between patients who will and won't respond to a particular sur-

gery or physical therapy treatment.

Currently, biomechanics datasets are growing, and big data methods to tackle those mountains of information are still under development. But there's enough of a track record to suggest the big data revolution can help injured runners get back on the road, keep people moving to fight obesity, and transform how physicians treat patients with movement disorders, such as cerebral palsy.

It will take hard work to overcome some of the challenges of biomechanics big data. More researchers need to share the data they have, some of which may be stashed on hard drives of individual graduate students. The heterogeneous nature of the data will require analytical tools that don't yet exist. Then there's the enormous and looming opportunity to take advantage of data from wearable sensors that track movement patterns of millions of users each day. But if these challenges can be overcome, the river of data could offer tremendous insight about human movement.

Big Data Challenge #1: Amassing Data

Ferber is not the only scientist who has longed for more research subjects. Most standard mobility studies involve just 10 to 100 people—hardly enough to populate a database for statistically meaningful comparisons. "To have over a million subjects is completely unheard of," says **Scott Delp**, **PhD**, professor of bioengineering at Stanford who is leading the University's effort in mobility data analysis.

To address that problem, Delp and his colleagues have established a consortium of four major clinical centers with mobility data from more than 20,000 people. "Over the last 10 years, we've been changing the culture of motion analysis labs. They feel comfortable and safe sharing their data," Delp says. He plans to expand the repository to include biomechanics data for over 10 million people.

Outside the controlled lab setting, **Ray Browning**, **PhD**, director of the Physical Activity Energetics/Mechanics Lab at Colorado State University, is four years into a five-year NIH-funded project that is accruing a massive quantity of data using sensors worn during what's called "free-living physical activity." His team is trying to determine if playground renovations and activity-promoting recess curricula encourage kids to be more active. They've fitted Denver area schoolchildren with wrist accelerometers that track their movements 75 times per second over a six-day period. Thus far, the researchers have gathered seven to eight gigabytes of mobility data from about 2,000 children—totaling in the terabyte range. "You could never process one of these files on a personal computer because there isn't enough memory to load it," Browning says.

In addition to big data collected by labs and free-living researchers, the field of biomechanics stands to benefit from a new frontier—mobility data captured daily by millions of people using wearable sensors such as FitBit and Jawbone as well as smartphone

health apps such as Azumio's Argus. "Even if you pooled data from all the gait analysis labs in the world, it wouldn't be as big as the database that FitBit already has," says Browning. "People are wearing these devices. I don't have to recruit them. It removes a huge hurdle."

Big Data Challenge #2: Heterogeneous Data

With all this data in the pipeline, biomechanics researchers have to address one of the hallmark challenges of big data: its diversity.

Calgary Running Clinic developed a 3-D gait analysis system to gather motion capture data. It is gathering big data not only at the Calgary clinic but also at about 30 clinics and universities around the world, according to 3dgait-analysis.com. Courtesy of Reed Ferber.

In studies of gait abnormalities due to cerebral palsy (CP)—the most common movement disorder in children—researchers gather multitudinous data from patients: They stud their arms and legs with reflective spheres that light up high-resolution images captured by video cameras, use electrodes to track electricity patterns in their muscles, and attach sensors to measure forces generated against the ground each time the child takes a step. In the end, doctors scrutinize video clips and pore through several years' worth of graphs and charts to make their best guess as to how the patient might fare given a particular surgery.

It would be better, Delp says, if researchers could take all that data and compare it to a database of similar information for other patients. They could then identify which surgical treatment worked best for people whose muscle patterns and motor activities resemble those of the patient at hand.

Recent research suggests the challenges of integrating these multifarious data types are not insurmountable. Stanford mechanical engineer **Jennifer Hicks, PhD**, and colleagues have created biomechanical models that make useful predictions by integrating several different types of data—muscle strength measurements collected during physical exams and biomechanical data from motion-capture computer simulations. When applied to a group of children who had trouble walking due to cerebral palsy, the algorithms identify which patients will benefit from surgery 70 to 80 percent of the time. "It's not great—you'd like to be at 95 percent—but definitely better than human experts," says Delp. Research led by **Mike Schwartz, PhD**, associate professor of orthopedic surgery at the University of Minnesota—and reported this issue in *Computation in the Surgical Suite*—also suggests that modeling and simulation can help predict surgical outcomes.

Chris Re, PhD, an assistant professor of computer science at Stanford, and colleagues are creating computer algorithms that integrate heterogeneous biomedical data from motion-capture video, smart phones and other sensing platforms. They are borrowing strategies they've already applied in paleobiology, where they designed a machine-reading system that amasses and integrates fossil data more effectively than databases manually compiled by experts.

The onslaught of data from wearable sensors and smartphones adds yet another dimension, and collaborations are beginning to emerge between companies producing these devices and researchers who use high-resolution imaging and motion capture. For example, Stanford computer scientists are working with researchers at Palo Alto–based Azumio to extract biomechanics data from the company's mobile health app, Argus. Tracking 500 million steps per day, as well as heart rate and sleep patterns, Argus is "like FitBit on your smartphone," says **Bojan Bostjancic, PhD**, Azumio's CEO and founder.

Granted, smartphone recordings are noisier than high-resolution video footage from lab cameras. But "it's extremely cheap data from millions of people 24/7," says data mining expert **Jure Leskovec, PhD**, assistant professor of computer science at Stanford and a Delp collaborator. "We can connect these types of data so they complement each other. That is our aim—to take high-quality data from labs and integrate it with large-scale noisier data from cell phones."

Big Data Challenge #3: Extracting Insight

Once big data is in hand and integrated, the task of learning from it can begin, as it already has. In a recent study, Ferber and his team compared the gait patterns of 34 people with osteoarthritis to a large dataset of gait information and treatment outcomes for others with the disease and other musculoskeletal injuries. This big data approach allowed them to predict, with 94 percent accuracy, which participants would benefit from a six-week exercise program. As they forecast, nearly a third did not improve from the strength training, suggesting alternative treatments should be found. Ferber's team will present its findings this July at the World Congress of Biomechanics in Boston.

The gait analysis database has also helped physical therapists give better advice to injured runners. By comparing clients' gait data to that of runners in the database with a similar demographic profile (e.g., male elite runner), experts can not only confirm someone has patella-femoral syndrome (aka "runner's knee") but also tell them it's because their hip abductors are weak, for example. "The data help us go beyond relieving symptoms. They direct us to the source of the problem," says **Blair Shular**, head physiotherapist at Glen Sather Clinic in Edmonton, Canada, where 156 patients have undergone 3-D gait analysis since the clinic purchased the $25,000 system in early 2011.

Making use of movement data uploaded into the cloud by millions of people using wearable sensors and smartphone health apps will require much more powerful analysis tools, Browning says. But the payoffs could be staggering. He and others envision using wearable sensor data to encourage healthy physical activity in people at risk for obesity, or to warn runners of impending injury.

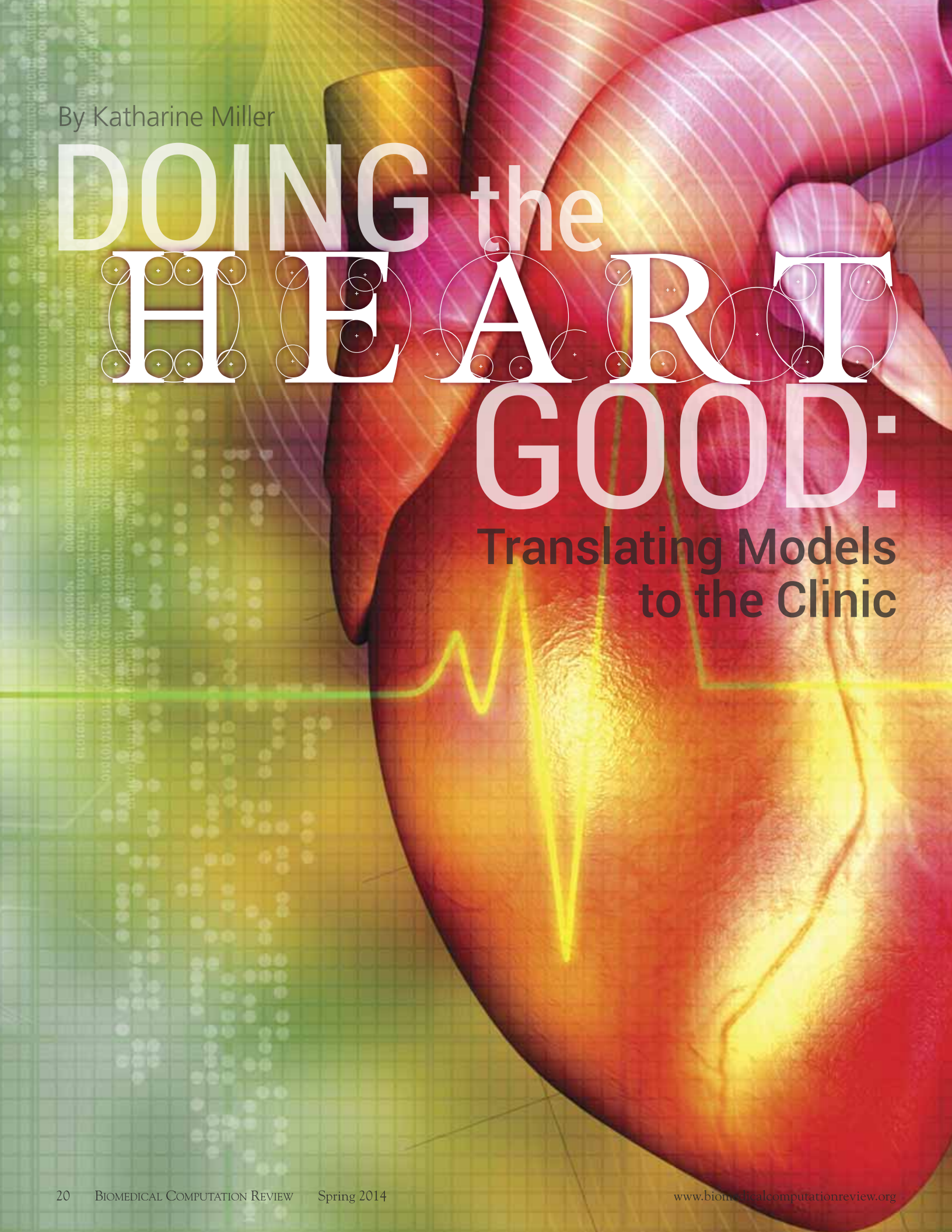
Many runners won't come in for gait analysis until they are hurt, Ferber notes. "We need to be able to tag them like wildlife, monitor them in their natural habitat, and analyze that data in a robust way to offer simple information like 'you're doing fine' or 'you're on the verge' or 'you'd better stop,'" he says.

Likewise, in studies of movement disorders, large datasets from wearable sensors might allow scientists to spot a pattern of shifts in center of mass just before people fall, and then use that insight for further research into fall prevention, Browning says.

And Delp thinks that large-scale, unprocessed mobility data from Argus and other wearable technology could help scientists understand the root causes of knee osteoarthritis and joint pain. "Instead of coming into the motion analysis lab once a year, they'd wear the unit every day for a year," he says.

One thing's for sure: Biomechanics researchers no longer have to stand in the rain to gather data. It's coming at them in a deluge. And they are ready for it. □

Fitbit, a wearable sensor (below), and Argus (right), a mobile phone app, are gathering data from millions of people. Biomechanics researchers are planning to tap into that treasure trove.



By Katharine Miller

DOING the HEART GOOD:

Translating Models to the Clinic

Pacemakers, heart-lung machines, stents and artificial hearts all sprang from successful partnerships between engineers and doctors. "From its inception in the 1950s, heart surgery has motivated advances in engineering technology, and engineering technology has enabled advances in medical treatment," says **Alison Marsden, PhD**, associate professor of mechanical and aerospace engineering at the University of California, San Diego. "I personally view modeling and simulation as the next step in this long history."

Modeling and simulation have the potential to help physicians answer some of the most fundamental clinical questions: Who is at risk if not treated? Who will benefit from treatment? And what is the best treatment for the patient in front of me?

Such clinical applications will only succeed if modelers and clinicians work closely together, says **Tain-Yen Hsia, MD**, a consultant cardiothoracic surgeon at Great Ormond Street Hospital in London, UK, who has collaborated with Marsden on modeling projects. Engineers need to educate physicians and surgeons about the potential and limitations of modeling. Clinicians, in turn, must educate engineers about the most clinically relevant questions and problems, he says.

Currently, only a few cardiovascular mod-

eling efforts are sufficiently lightweight for actual use in the clinic, but that is rapidly changing. And there is one sad fact that works in modelers' favor: Too many people are either being overtreated or undertreated because doctors don't have enough information to accurately predict risks and outcomes. Models have the potential to change that—and they only have to do better than current practice, a lower bar than one would think.

WHO IS AT RISK IF NOT TREATED?

Modeling and simulation can be used to noninvasively stratify patients' risk of various cardiovascular conditions (such as heart attack, arrhythmia, atherosclerosis, or blood clots). This can help to reduce overtreatment of lower-risk patients and undertreatment of higher risk patients, Marsden says.

Modeling Heart Disease Risk

One company, HeartFlow, Inc., seems to have hit the sweet spot. Their service seeks to improve how clinicians assess a person's risk of heart disease due to reduced blood flow in the coronary artery. Currently, the process often goes like this: If imaging shows more than 50 percent narrowing of the coronary artery (a condition known as stenosis), cardiologists send patients to the cardiac catheterization lab (cath lab) to have a stent implanted. Evidence suggests, however, that anatomical narrowing is not a good proxy for reduced blood flow, says **Charles Taylor, PhD**, founder and chief technology officer at Heartflow. A better measure is called the fractional flow reserve (FFR). Here's how it works: In the cath lab, the interventional cardiologist administers a drug to make the heart pump as it would under exercise; inserts a wire into the artery to measure the blood flow upstream and downstream of the narrowed region; and calculates the ratio between these flows—the FFR. An FFR of less than 0.8 indicates reduced function and the need for a stent or other procedure to widen the artery.

It turns out, though, that only about 20 percent of patients who are sent to the cath lab with coronary artery narrowing have greater than 50 percent diameter reduction, according to a recent study published in the *Journal of the American College of Cardiology* (JACC) in 2012. Many of these lesions won't be functionally significant. In other words, 80 percent of patients who go to the cath lab for stenting or FFR assessment—expensive procedures that carry risks of their own—don't need to be there, Taylor says. "The artery may be anatomically narrowed but it may not really matter in that specific patient."

This is where HeartFlow comes in. Using CT images from individual patients, Taylor and his team can simulate blood flowing through the coronary artery under exercise conditions and then calculate the FFR noninvasively. In a recent study of 254 patients, this personalized modeling approach, termed FFR_{CT}, proved capable of reclassifying 68 percent of the false positives as negative. "FFR_{CT} may spare many patients a trip to the cath lab," Taylor says. The work was published in January 2014 in the *JACC*.

HeartFlow's technology is approved for use in Europe and is currently waiting for FDA clearance in the United States. Taylor notes that HeartFlow's product is unusual because it doesn't try to predict an outcome. "We just need to show how good our measurement is," he says. So far, Taylor is quite confident that the procedure is better than the current protocol, which is dominated by what many cardiologists jokingly call the "occulo-stenotic reflex": See stenosis and treat it.

> Heartflow seems to have hit the sweet spot with a service that seeks to improve how clinicians assess heart disease risk due to reduced blood flow in the coronary artery.

Modeling the Risk of Losing Heart Rhythm

Natalia Trayanova, PhD, the Murray B. Sachs professor of biomedical engineering and medicine at Johns Hopkins University and Chief Scientific Officer at Cardiosolv LLC, is developing patient-specific computer heart models to predict which patients are at risk of dangerous arrhythmias—ventricular tachycardia [VT] or ventricular fibrillation [VF]—due to scarring from a heart attack.

Currently, to assess the risk of VT or VF, cardiologists measure a heart attack patient's left ventricular ejection fraction—the percentage of blood in the ventricle that is pumped out with each heartbeat. If a patient's ejection fraction is less than 35 percent (normal is 55 to 70 percent), the physician implants an ICD (implantable cardioverter defibrillator), a special kind of pacemaker that automatically detects heart rhythm and applies a shock to bring the heart back into rhythm. Yet, annually, fewer than five percent of these implanted ICDs ever have to fire to terminate an arrhythmia, Trayanova says. Thus, many people who were implanted don't need the device; and on the flip side, some heart attack patients with ejection fractions greater than 35 percent—who don't receive the device—need it, she says.

Trayanova wants to see if modeling can do a better job of predicting who is actually at risk of a dangerous arrhythmia. To that end, her team has simulated electrical activity in 40 patient-specific heart models of people who have ICDs. The result: In cases where the model predicted no arrhythmia, the devices have never fired.

The team still has to determine the minimal set of simulations that will be optimal for determining risk and to streamline the process. But, she notes, "We just have to do better than the current protocol." And since the protocol leads to the costly implantation of ICDs in patients who don't need them and non-implantation in people who do need them, Trayanova's models have the potential to improve patient care while also reducing costs.

Modeling the Risk of a Clog

The Heartflow FFR_{CT} protocol evaluates impaired coronary flow reserve at the time of examination. It does not, however, attempt to predict the future—for example, whether a patient is likely to develop atherosclerosis—a condition in which the walls of the arteries thicken and harden with plaques. The ability to predict the progression of plaque would be extremely useful for determining who should receive medical treatments such as statins.

Currently, physicians review MRI or CT scans and blood test results to try to infer

Translating Models to the Clinic:

Advice to Engineers

Collaborate deeply.

It used to be that clinicians were deeply skeptical of working with computer modelers. That was at least in part because modelers tended to solve the most numerically interesting problems rather than those with the greatest potential impact on medicine, Marsden says. But these days, "Getting collaborators is usually the easy part," she says. "The challenge is narrowing down the key clinical questions we can contribute to effectively." To do that, she says, engineers cannot work in isolation. "You have to have constant interaction with the clinical team so you can improve the model, incorporate new data, and change the question as you start to get answers."

Keep it clinical.

"One of our translation goals is to get the models to produce quantities the clinicians are familiar and comfortable with," Marsden says. Her computational fluid dynamics (CFD) models provide information on local hemodynamics and energy loss—terms rarely used by physicians. So she wants to couple the CFD models to physiological models that will allow calculation of measurements of greater interest to clinicians, such as oxygen saturation, pressure loss, or cardiac workload.

Similarly, Trayanova notes that if a model can reproduce quantities with which clinicians are familiar—electrocardiogram results, for example—"then they are more likely to believe."

Make it cost-effective.

Economic modeling of the cost-effectiveness of FFR_{CT} suggested that it would save money while having no adverse impact on patient outcomes—and perhaps even having a positive one. The work was published in *Clinical Cardiology* in 2013. In a multisite study, Heartflow is now comparing the economic and clinical outcomes of patients evaluated with either FFR_{CT} or the current standard of care. Taylor believes that the results will also show cost savings with no adverse outcomes. By doing better than the current protocol—and doing it efficiently and noninvasively—Heartflow has received positive feedback from many health insurance providers, Taylor says.

Remember the FDA.

A critical and sometimes challenging part of any effort to bring models to the clinic is the need for FDA approval or clearance. Heartflow is currently waiting to hear back from the FDA on its application for clearance to sell its service, at which point the company should become eligible for insurance reimbursement.

For some models, it's unclear whether FDA review will be necessary. For example, Marsden's Y-graft for the Fontan procedure wouldn't require FDA approval if seen as a surgical procedure (because a surgeon is free to implement whatever design s/he thinks is best for the patient), but would require approval if deemed a device. "The Y-graft involves sewing together a trunk with two branches. Whether that's a device or surgical method is perhaps fuzzy," she says. It gets even fuzzier if the grafts are made patient-specific (which Marsden believes would improve outcomes).

The FDA is also still figuring out exactly what data they will require to demonstrate that a simulation tool is sufficiently reliable for clinical use. They will likely want to see an error bar or confidence interval before deciding if something is safe for the clinic, Marsden says. So modelers need to start doing uncertainty analyses. "Clinical data is full of uncertainty, and you need to know the uncertainties of the input and how those propagate through to the model output." Since uncertainty quantification is a field unto itself, Marsden advises teaming up with people with that expertise.

Make it lightweight.

"There's always a tension between doing things that are cool and doing things that are practical," says **Hiroshi Ashikaga**, **MD**, **PhD**, assistant professor of medicine and biomedical engineering at Johns Hopkins University. "But the ultimate test is whether clinicians would like to use a model." Portability and simplicity are key. "Clinicians want to get things done." So modelers should try to develop portable systems where all the clinician has to do is download images, put it in the model, and get the results, he says.

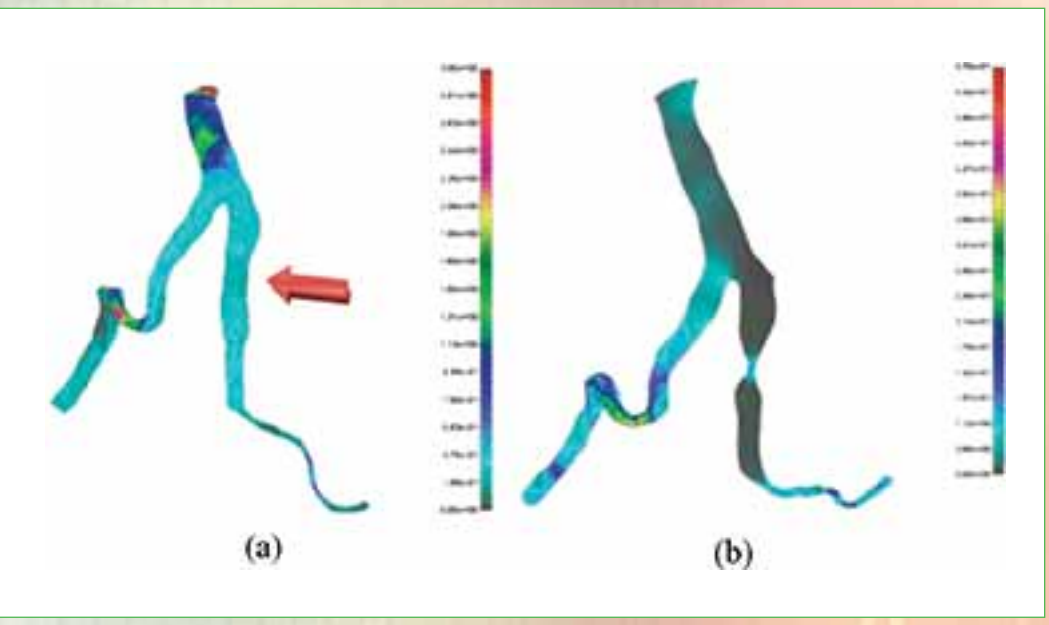
To reduce computational times, researchers are trying time-saving computational approaches, such

how a patient's clogged arteries will evolve: Will they get more clogged or stay the same? And how fast will they change? "Despite being an experienced clinician, I couldn't do this in my head," says **Oberdan Parodi, MD**, a cardiologist at the Ospedale Niguarda Ca' Granda in Milan, Italy.

So Parodi collaborated with researchers who were part of Europe's Virtual Physiological Human project ARTreat to develop a multi-scale computer model of plaque formation and evolution. It combines 3-D image reconstruction of the arteries with blood-flow modeling and models of the initiation and progression of plaque, as well as plaque characterization.

Parodi's model considers the transport and chemical interaction of LDL and HDL; the role of adhesion molecules; the movement of cells and other materials in the bloodstream; and arterial wall thickening. In the end, the model calculates wall shear stress—the forces exerted on the arterial walls—using computational flow dynamics. And by calculating it at multiple points along the artery wall, Parodi can detect the locations where plaque will likely form. "That's relevant from a prognostic point of view for managing this patient," he says.

Parodi validated his model in pigs (who could be sacrificed) but thus far has only been able to test the model in 64 relatively healthy human patients who had less than 50 percent plaque blockage (patients with major blockages need to be treated immediately). "We were able to see several plaques that progressed as well as the new onset of plaques over time and in places the model predicted," Parodi says. More human studies are needed, he says. "This model is still a prototype. Getting it into the clinic requires more testing and money."



Parodi modeled plaque progression in the coronary arteries of 40 patients over the course of two years. The colors indicate arterial wall stress, with gray representing low stress. Here, the models observed an increase in plaque and a narrowing of the vessel after six months (b) at the original site of lowest wall stress (red arrow in (a)). Reprinted with permission from Parodi, O,* et al.*, Patient-Specific Prediction of Coronary Plaque Growth from CTA Angiography: A Multiscale Model for Plaque Formation and Progression,* IEEE Transactions on Information Technology in Biomedicine *(2012).

WHO WILL BENEFIT FROM TREATMENT?

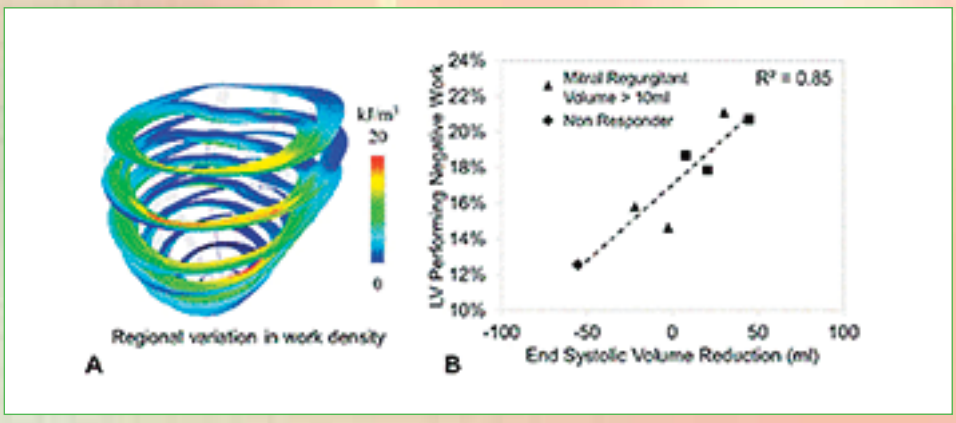
Some models go a step further than the models described above: They simulate a possible treatment and try to predict the outcome. The goal is the same—to identify which patients will likely benefit from treatment—but the models simulate treatments in addition to patient physiology.

Modeling Who to Pace

About 40 percent of patients with heart failure suffer from a double whammy: They have defects in both electrical conduction and pumping. "Their contractions are not only weaker, but also less synchronous," says **Andrew McCulloch, PhD**, professor of bioengineering at UCSD. These patients are typically implanted with two pacemaker leads—one on each ventricle—to resynchronize the contractions between the left and right sides of the heart. When this cardiac resynchronization therapy (CRT) works, it improves the mechanical function of the heart.

For unknown reasons, CRT fails about one-third of the time. But McCulloch's team may be the first to be able to predict when this will happen. Using patient-specific modeling in eight patients, his team found a nearly linear correlation between successful CRT and a calculable quantity called "heterogeneity of regional work." Essentially, CRT is most beneficial when the distribution of work across the heart muscle is not uniform.

McCulloch's team plans to test the model in more patients. But here's the catch for translating this model to the clinic: You can't measure regional work directly; you have to calculate it based on detailed measurements—some of which are invasive and not routine prior to CRT. Nevertheless, McCulloch is optimistic: "I



McCulloch's detailed models of hearts with dyssynchronous heart failure revealed a close correlation between heterogeneous regional work (below left), which was highly variable among patients, and successful cardiac resynchronization therapy (CRT). In addition, there was a nearly linear correlation between the portion of the left ventricle performing negative work and CRT success (below right). Courtesy of Andrew McCulloch.

actually think we will be able to get the same result with less invasive measurements and use invasive measurements to validate the model," he says. And at that point perhaps this approach will be ready for development into a viable clinical application.

Modeling Who to Graft

If a patient's coronary artery becomes severely blocked with plaque, flow can be improved by bypassing the clogged area with a grafted artery or vein obtained from somewhere else in the patient's body. But vein grafts used for this procedure (coronary artery bypass grafting or CABG) fail at alarmingly high rates, says Marsden, and more often than artery grafts.

When Marsden first began approaching this problem, she planned to model how to optimize graft geometries and placements for specific patients. But clinicians told her that wasn't the challenge; They just wanted to know why certain patients' grafts fail so often. Marsden listened and her team switched gears. In collaboration with Jay Humphrey at Yale University, her group began simulating how veins remodel under changing flow and pressure conditions and then paired those simulations with patient-specific models of post-CABG patients.

Vein grafts are known to experience a big shock when implanted in the arterial system: They are suddenly exposed to a 20-fold increase in pressure. Marsden's team's simulations showed that veins appear to adapt more favorably when the load is applied more gradually. The work is still relatively new, but if it leads to further insights that can be validated in the lab, Marsden says, "ultimately they could result in a clinical solution."

WHAT IS THE BEST TREATMENT FOR THIS PATIENT?

Perhaps the most challenging type of translational modeling and simulation involves looking for better ways to treat patients. In some cases, the status quo is a pretty unpleasant option—so improvements are desperately needed.

Modeling Where to Burn

In addition to modeling the risk of ventricular tachycardia (VT), Trayanova has

Advice, Continued

as automated mesh generation or the use of beam elements in their finite element models. When trying to reduce computational cost, Capelli says, researchers need to determine when good is good enough. "But good enough depends on the question we're addressing," he says. If it's a question of whether a device fits or not, then the models can compromise on numerical accuracy in favor of reducing computational times. But if the model is exploring the risk of structural failure of a device, it requires more precise information, he says.

Commodify it.

In any effort to translate research to the clinic, there is a "valley of death" to be overcome, says **Ahmet Erdemir**, **PhD**, assistant staff in the Department of Biomedical Engineering at the Lerner Research Institute of the Cleveland Clinic. Business savvy and venture funds may be required to take a research-worthy model and make it clinic-worthy, he says.

Heartflow (Taylor), Cardiosolv (Trayanova) and InsilicoMed (McCulloch) are all examples of computer modelers bringing a business mindset to their computational models. Erdemir also has a business-like establishment at the Cleveland Clinic paired to his academic work—Computational Biomodeling (CoBi) Core. His academic research program generates some tools that this fee-for-service facility can use, and CoBi Core generates new ideas that he can feed back into the research program. "My NIH program officer referred to it as my own P41," he says.

Merge with the existing workflow.

For computer models to move into the clinic, they must be useful without greatly disrupting clinical practice. Heartflow's FFR_{CT} is a case in point: It doesn't require any extra protocol in the clinic—not even an extra exam, Taylor says. When the patient's CT is completed, a virtual machine at the clinical site securely pushes anonymized images through a firewall to Heartflow. The results are then automatically sent back to the physician. It's extremely non-disruptive and can be easily integrated into clinical practice.

Parodi has a different approach: he is building his models of plaque composition and progression into a clinical decision support system called ARTreat that can be deployed at the clinic. The entire system is not yet ready for prime time, although some pieces of it—including software for automatic analysis of plaque components from noninvasive imaging—have been tested by cardiologists with positive results.

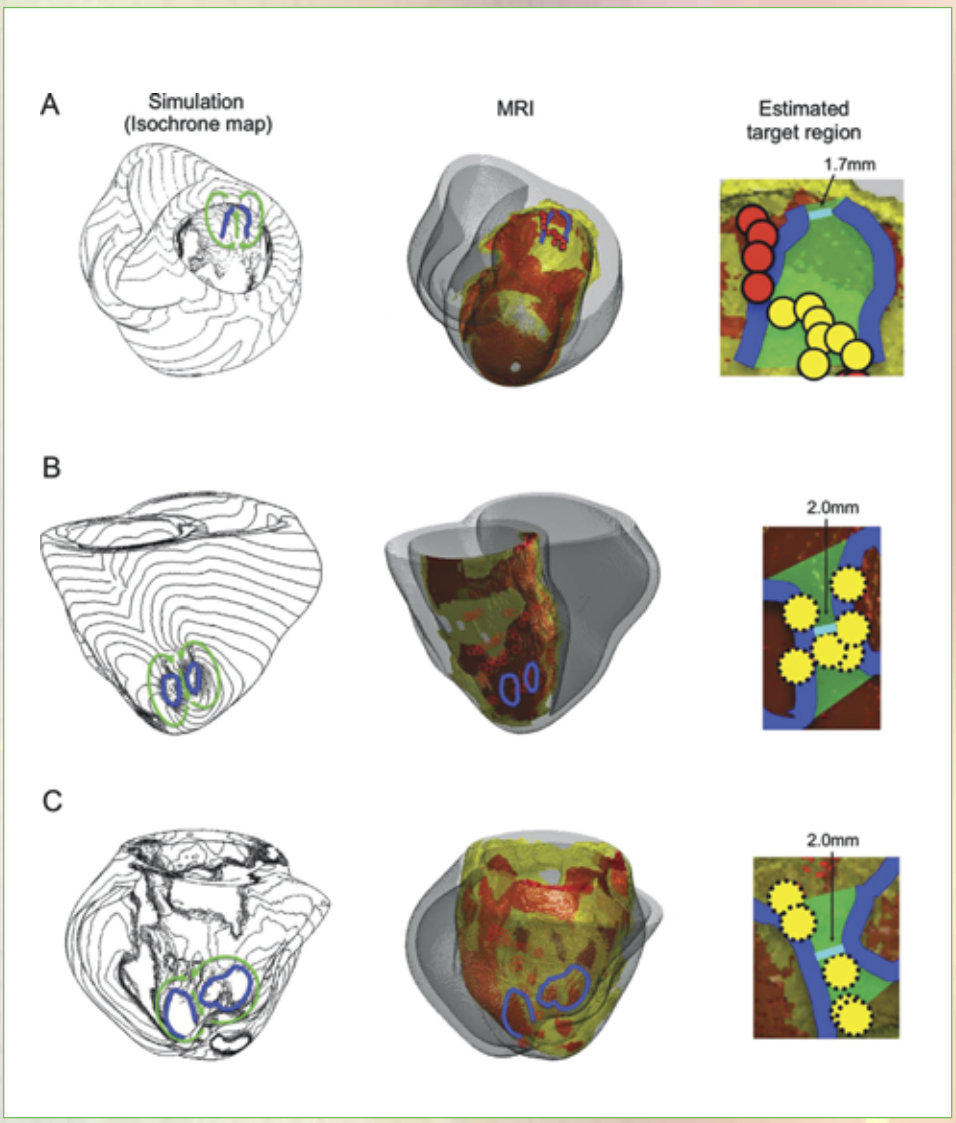
Some models might reach the clinic almost invisibly. For example, if Marsden's Y-graft procedure proves beneficial, it could become widely adopted as a surgical method without further reliance on the model. Similarly, McCulloch's models that determine who will respond to CRT might lead to clinical criteria for deciding who to treat, he says, "possibly obviating the need for patient specific modeling."

Such a result might not lead to money or fame, but it will make a difference in the clinic. And that's what it's all about. □

been modeling the best way to perform a common treatment called ablation—inserting a probe to selectively burn heart tissue as a means of disrupting the circuit that causes VT. It's a brutal six to eight hour procedure: The interventional cardiologist inserts a probe in the heart to cause VT; makes an educated guess as to where to burn to prevent VT; does the ablation; and then shocks the heart back into rhythm. It is then necessary to repeat that aggressive procedure several times until VT can no longer be provoked. Along the way, the electrophysiologist also makes many burns.

In hopes of shortening the procedure, Trayanova and her team have been studying the feasibility of using modeling to predict where a surgeon should ablate. In a paper published in *Heart Rhythm* in 2013, they reported on work in which they performed 30 to 50 simulations of electrical conduction on 13 patient-specific models of VT. It required a lot of computer time, but it was time well spent, Trayanova says. The models blindly predicted ablation sites that the electrophysiologists did in fact use. The team is now working on a follow-up paper reporting on simulations with models of more patients' hearts and demonstrating that if electrophysiologists relied on the simulations rather than guesswork, ablation procedures would be shorter and require fewer burns, Trayanova says.

CardioSolv LLC, the company founded by Trayanova, has been trying to build a more lightweight version of the model, as well as find a way to streamline the delivery of model predictions to the electrophysiologist. "We're scaling back to figure out what are the most important components," Trayanova says.



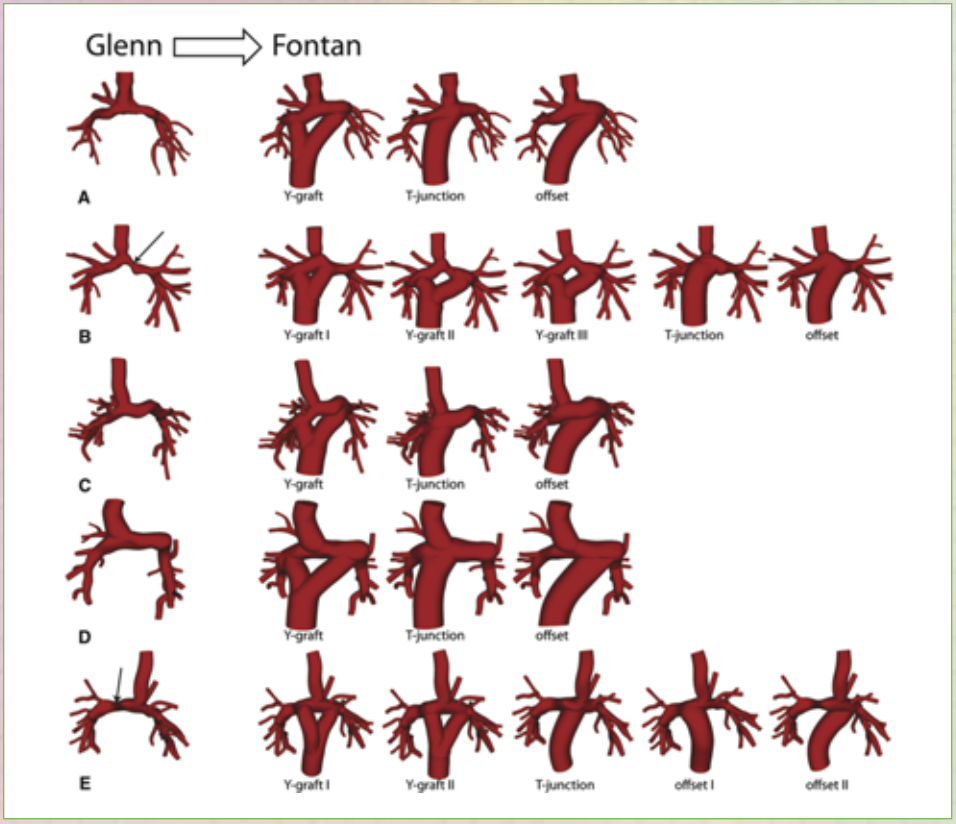
When Trayanova's team simulated ventricular tachycardia (first column) in patients A, B, and C, they identified areas where conduction was blocked (blue lines). In the magnetic resonance imaging (MRI) column, the blue lines from the image-based simulation are co-registered with the actual ablation sites (red circles) on the heart geometry. The last column shows a zoomed-in image of the zone (green) that the simulations predicted as a likely target for successful ablation, along with a cyan line marking the shortest possible line of ablation across the target region. The ablation sites that fell within the estimated ablation target (green area) are indicated by yellow circles and are consistent with the simulations' predictions that ablations in this target zone should be sufficient to prevent VT. Reprinted from Hiroshi Ashikaga,* et al.*, Feasibility of image-based simulation to estimate ablation target in human ventricular arrhythmia,* Heart Rhythm *10/8:1109-1116 (2013), with permission from Elsevier.

Modeling Surgical Corrections

Some children are born with only one ventricle in the heart—a dangerous condition that is fatal if left untreated. Typically, cardiologists correct the problem by performing three open-heart surgeries, culminating in what's known as the Fontan surgery: They insert a Gore-Tex graft in various conformations (typically a T-junction or an offset T) to redirect and improve blood flow.

Because surgeons wondered if the existing grafts were optimal, Marsden and her colleagues decided to model and simulate the procedure using several alternative graft shapes. The simulations led to a proposed Y-graft modification of the procedure that appeared to improve blood flow and energy efficiency. A pilot study in six patients found that the surgery is technically feasible and has outcomes that are at least similar to the standard of care, Marsden says.

While the Fontan procedure saves lives, many patients can still develop long-term complications such as pulmonary arteriovenous malformations, Marsden says. Long-term follow up data on the Y-graft is therefore needed, Marsden says. "We may



Marsden and her colleagues modeled a novel Y-graft version of the Fontan procedure, which corrects for a congenital heart defect in which babies are born with only one ventricle. This diagram displays multiple different surgical design options in five different patient anatomies (A to E). Computational fluid dynamics simulations predicted the Y-graft would lead to improved blood flow and reduce energy losses. Reprinted with permission from Weiguang Yang,* et al.*, Hepatic blood flow distribution and performance in conventional and novel Y-graft Fontan geometries: A case series computational fluid dynamics study,* J Thorac & Cardiovasc Surg *143/5:1086-1097 (2012).

not find out if it improves outcomes for another 5 to 10 years."

Modeling the Device/Patient Matchup

While tight working relationships between modelers and clinicians are essential to any kind of biomedical modeling for the clinic, they may be even more important when modelers work at the interface between medical device engineers and clinicians.

Melissa Young, PhD, project scientist in the biomedical engineering department of the Cleveland Clinic and assistant professor at the Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, has been working closely with a multidisciplinary team that is developing a mitral valve frame—a kind of valve stent—that can be delivered noninvasively. The team brings together modelers, surgeons, imaging experts, device designers, and testing engineers, not to mention the Cleveland Clinic's innovations group, an entity that seeks industry sponsors and other funding for projects such as hers.

Because the mitral valve frame will be delivered by a catheter, it has to start small and then expand into place. It therefore needs to be both strong and flexible while also having a good fit with the mitral valve anatomy. Using simulations of various stent designs and mitral heart valve boundary and loading conditions, Young's team identified a few favorite designs out of several dozen possibilities. "Manufacturing stents for evaluation is expensive," Young says. "If you can take 25 designs and narrow them down to three that would be best for fatigue, gross expansion, and the ability to anchor successfully, that's really helpful."

Young is also developing a computer tool to guide surgeons in selecting stents for their patients with peripheral artery disease, a condition that affects about eight million Americans and carries health care expenditures of $3.7 billion annually. Currently, surgeons have a choice of just a few stents to help these patients; and they might tend to select the ones they are used to—or what the hospital carries. "But that's not necessarily the best fit for a patient," Young says.

When completed, Young's tool will allow surgeons to try different devices on a virtual map of the patient's arteries. This map, which is a mesh inferred from patient-specific intravascular ultrasound data, shows not only arterial structure but also plaque composition, a characteristic that affects arterial mechanics and, in turn, drives stent choice or placement.

Claudio Capelli, PhD, research associate in the Centre for Cardiovascular Imaging at University College London Institute of Cardiovascular Science, also works on matching devices to patients. He has been developing patient-specific models to explore percutaneous pulmonary valve implantation (a procedure for replacing the pulmonary heart valve using a catheter). The procedure is extremely beneficial but still not widespread, Capelli says.

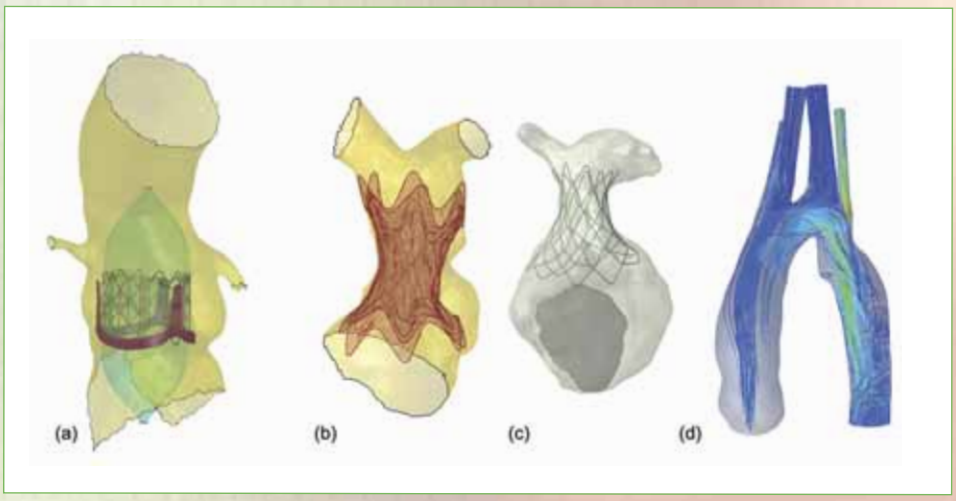
Capelli works in a team that was established by **Silvia Schievano, PhD**, a decade ago and is located within a pediatric hospital where the engineers participate in clinical meetings for non-routine, difficult cases. "We offer the possibility of simulating an intervention before it happens, especially when there is doubt about which device to use or which size to use because of interaction with the anatomy," he says.

Starting from images, his team constructs anatomical models of a patient's heart, which might take two or three hours, he says. Next, they simulate the implantation of a handful of devices that might be appropriate based on the cardiologist's in-

"Manufacturing stents for evaluation is expensive," Young says. "If you can take 25 designs and narrow them down to three that would be best for fatigue, gross expansion, and the ability to anchor successfully, that's really helpful."

dications. The simulations then determine the stresses on the device and heart. Finally, the team offers its insights to the cardiologist who makes his or her own decision about what to do. "Whether they agree with our prediction or not," he says, "it still provides us a huge source of data to follow up and validate our predictions."

Thus far, Capelli's team has provided this kind of predictive advice in about ten patient cases. In one case, his team was asked to determine the optimum stent size to correct a narrowed aorta with very unusual anatomy. "We needed to cover an aneurysm without covering access to a small vessel on the other side," he says. His team did a 3-D reconstruction of the anatomy and used finite element analysis and computational fluid dynamics to recommend a particular device. The cardiologist followed the team's suggestion and the procedure was successful, Capelli says.



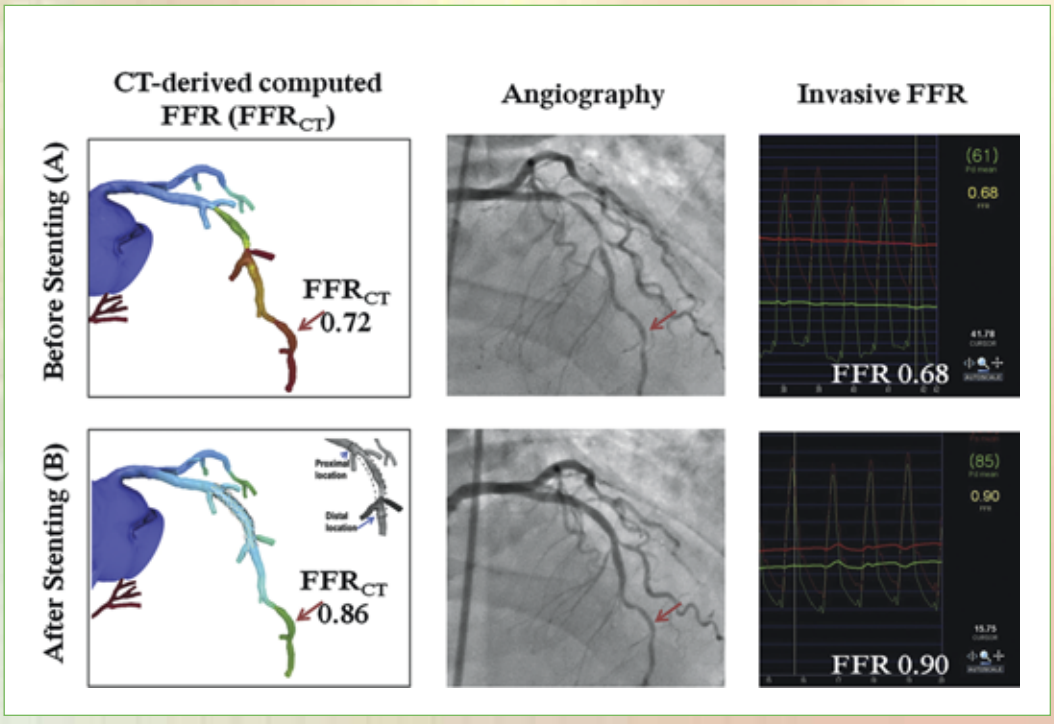
Capelli's group has performed patient-specific simulations for a variety of interventional procedures including (a) transcatheter aortic valve implantation: case of valve-in-valve procedure; (b) percutaneous pulmonary valve implantation: stent-graft; (c) percutaneous pulmonary valve implantation: bare metal stent; (d) flow distribution following the implantation of a stent-graft within aortic coarctation. (Simulations by C Capelli and GM Bosi).

Sometimes modeling also gets it wrong. In one case, Capelli's group selected a device with a certain size (the largest available). But when the cardiologist did a balloon sizing (to evaluate the location), he found a less resistant wall than what the model predicted. "The wall didn't guarantee a good anchoring," Capelli explains. Since the recommended device was already the biggest available, no device was implanted. "So the cardiologist always has the last word," he says.

Capelli and his colleagues continue to refine their models for device choice, and have also developed a virtual library of 41 different devices and 600 to 700 models, most of them related to congenital heart disease. Capelli hopes device-development companies will use these libraries to test various devices on particular target populations. "Having the possibility to play virtually with these anatomies can help reduce animal tests and replicate more realistic conditions in the testing of devices," he says.

Virtual Stenting

HeartFlow is also entering the realm of simulating alternative treatments. To demonstrate his latest project, Virtual Stenting, Taylor pulls out his iPad and loads a 3-D image of a patient's heart and arteries. He then touches the narrowed sections of the artery to get a pop-up showing the FFR_{CT}. If the FFR_{CT} is less than 0.8, he selects a virtual stent from a menu, drags it into place, re-sizes it, and then touches the artery to see how the stent changes the FFR_{CT}. The software was successfully tested prospectively on 44 patients: FFR_{CT} for the patients closely matched the invasively measured FFR both before and after stenting. The work was published in *JACC: Cardiovascular Interventions* last year, and is perhaps the clearest example yet of how lightweight cardiovascular computer models could find their way into the practice of medicine. □



Heartflow is developing a virtual stenting application to evaluate whether a stent will change a patient's stenosis. Here, before stenting (top images), stenosis of the left anterior descending (LAD) coronary artery is demonstrated by a noninvasive FFR_{CT} of 0.72 and confirmed by invasive coronary angiography and invasive FFR (0.68). After virtual stenting (bottom images), noninvasive FFR_{CT} demonstrated no ischemia in the LAD, with a computed value of 0.86. Invasive FFR after stent implantation was 0.90. Reprinted from Kyung-Hee Kim, et al., A Novel Noninvasive Technology for Treatment Planning Using Virtual Coronary Stenting and Computed Tomography-Derived Computed Fractional Flow Reserve,* J Am Coll Cardiol Intv. *7(1): 72-78 (2014).

Under TheHood

BY GUY HASKIN FERNALD

Getting Started with Cloud Services for Biomedical Computation

Biomedical researchers who work with large data sets may run out of both disk space and patience while waiting for a computation to finish. Though buying more hard drives and faster computers may seem tempting, the cloud is now a realistic option.

In 2008, when cloud computing was relatively new, this magazine published a column by Alain Laederach predicting that scientists would be won over to cloud computing, despite some people's concerns about a loss in performance with the added layer of virtualization.

In the last six years the world of cloud computing has expanded dramatically. Moreover, the performance losses skeptics feared have not surfaced. I have used both cloud computing and local clusters and have found that after some initial setup, computing in the cloud can be just as efficient as working on a local cluster. And the ability to quickly change the number of machines in the cluster allows scaling up or down to fit the problem at hand.

> ...The ability to quickly change the number of machines in the cluster allows scaling up or down to fit the problem at hand.

For biomedical researchers, several realistic cloud options now exist, including Amazon Web Services, Joyent, Google's Compute Engine, the HPCloud, IBM SmartCloud, and Rackspace—all with very low barriers to adoption, including step-by-step tutorials and guides to setting up public key cryptography to permit secure access to a virtual server with Secure Shell (SSH).

Amazon Web Services (AWS), the most well known of the cloud computing services, launched in 2006 and now includes dozens of products. The most relevant for scientific cloud computing is the Elastic Cloud Compute (EC2) service, which provides access to a variety of virtual machines ranging from small (1 CPU, 1.7GB RAM for $0.60/hour) to very large (32 CPUs, 244GB RAM for $2.40/hour) as well as high-performance GPU machines.

AWS users can select an Amazon Machine Image (AMI), which contains the operating system and software for the virtual computer. There are several Linux-based AMIs that come pre-installed with bioinformatics related software, including the Bioconductor AMI and the CloudBioLinux AMI. Once an instance is running, users receive a public IP address which they can use to upload data and log in to run programs. Users who want more computing power can start up more servers or even create a virtual compute cluster with as many nodes as desired using the StarCluster program from MIT.

There is one major drawback to using virtual servers: Once the server is shut down, it ceases to exist, creating a risk of losing valuable data or results if they are not copied to a local computer before terminating a cloud instance. One solution: Amazon's Elastic Block Storage (EBS) volumes, which hold up to 1TB of data and can be mounted just like a disk drive. The user can also reattach a virtual machine to the EBS volume at a later time to continue working. Unfortunately EBS volumes can reduce performance due to slower disk read and write operations. The Manta service from Joyent has some potential advantages for high-performance computing on large datasets because it focuses on the data storage first and then brings the computation to the storage. By integrating the computation with the data, Manta avoids slow performing network drives and operates with virtually zero data latency.

Cloud computing can be a cost-effective and flexible solution for a researcher's computing needs. When tackling a large data analysis project, I recommend considering an upload to the cloud. □

Resources

Amazon Web Services:
http://aws.amazon.com/

BioConductor AMI:
http://www.bioconductor.org/help/bioconductor-cloud-ami/

CloudBioLinux AMI:
http://cloudbiolinux.org/

Joyent's Manta service:
http://www.joyent.com/products/manta

MIT's StartCluster program:
http://star.mit.edu/cluster/

DETAILS

Guy Haskin Fernald is a PhD candidate in Russ Altman's lab at Stanford University. He is working on identifying and using molecular features of drugs to predict chemical activities and biological phenotypes. Cloud computing is proving to be one of his best tools for working with large chemical databases and implementing machine learning algorithms.

DIVERSE DISCIPLINES, ONE COMMUNITY
BiomedicalComputation REVIEW
Published by Simbios, an NIH National Center for Biomedical Computing

Nonprofit Org.
U.S. Postage Paid
Permit No. 28
Palo Alto, CA

Stanford University
318 Campus Drive
Clark Center Room S271
Stanford, CA 94305-5444

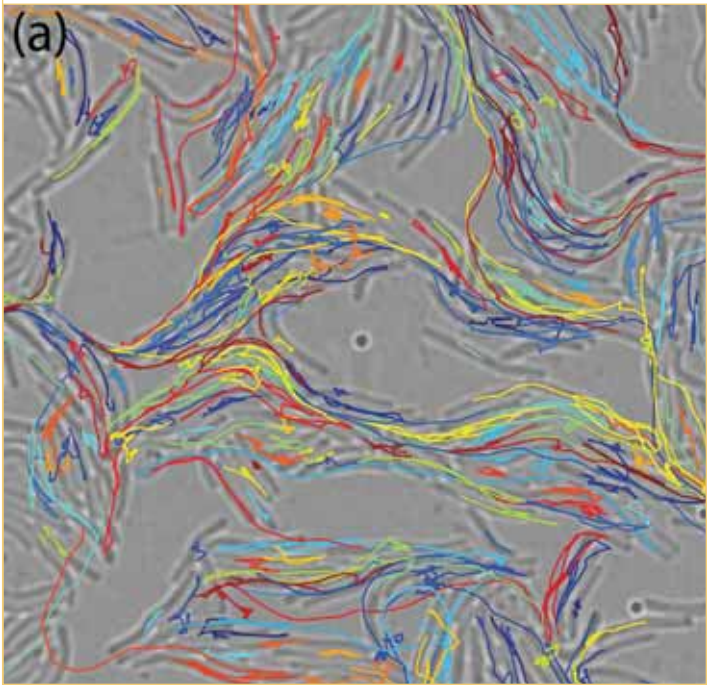
seeing science

SeeingScience

BY KATHARINE MILLER

Tracking Bugs

To better study the interactions and movements of whales and birds in the wild, researchers tag them with GPS tracking devices. But tracking critters in more crowded environments, such as a Petri dish, can be challenging—especially if, like fruitflies, they zig and zag in overlapping movements; or, like bacteria, they wriggle and glide while in contact with one another.



"If you have multiple objects far from each other, it's easy. But when they get closer, the tracker gets confused," says **Joshua Shaevitz, PhD**, associate professor of physics at Princeton University and the Lewis-Sigler Institute for Integrative Genomics.

To remedy that problem, Shaevitz and his colleagues built on an approach called an active contours or "snake" algorithm, which treats the bright parts of an image as a kind of energy that needs to be maximized. "You put wiggly lines around and they try to dock where the energy is," Shaevitz says. But, typically, the snakes are allowed to merge, break, fork, and recombine.

In Shaevitz' team's innovation, the algorithm knows how many objects are present and assumes not only that they don't disappear or get generated, but also that they repel each other—preventing them from collapsing into one another if they overlap.

The algorithm can track any objects that get close to each other, and is already widely used for tracking *Drosophila melanogaster*, the fruitfly commonly used for research. □

Shaevitz is using his team's tracking algorithm to follow the movements of* Myxococcus Xanthus*, a bacterium that glides like a long flexible worm and coordinates movements among large structured groups. Here, the trajectories of 205* Myxococcus Xanthus *cells over 2000 seconds are indicated in different colors, overlaid on the first frame of the movie. Using the data collected, Shaevitz and his team calculated each bacterium's gliding speed. Reprinted from Deng Y, Coen P, Sun M, Shaevitz JW (2013) Efficient Multiple Object Tracking Using Mutually Repulsive Active Membranes.* PLoS ONE *8(6): e65769. doi:10.1371/journal.pone.0065769.