DIVERSE DISCIPLINES, ONE COMMUNITY

Biomedical Computation REVIEW

Published by Simbios, an NIH National Center for Biomedical Computing

THE Epigenome

A New View Into the Book of Life

PLUS:

The Physiome:
A Mission Imperative

Summer 2010

contents

ContentsSummer 2010

FEATURES

8 The Physiome: A Mission Imperative
BY KATHARINE MILLER

16 The Epigenome: A New View Into the Book of Life
BY KRISTIN SAINANI, PhD

DEPARTMENTS

1 GUEST EDITORIAL | DEMOCRATIZING INTEGRATIVE BIOLOGY
BY ATUL BUTTE, PhD

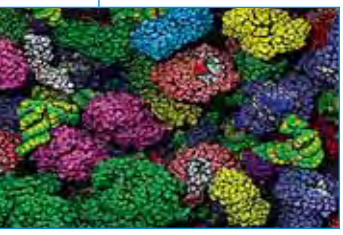
2 SIMBIOS NEWS | REACHING UNDER THE HOOD OF A 20-YEAR-OLD MUSCULOSKELETAL MODEL BY JOY P. KU, PhD

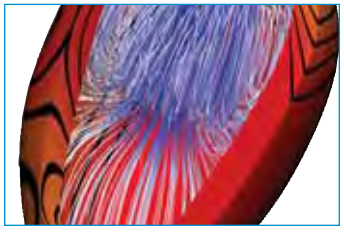
3 NEWSBYTES | BY LOUISA DALTON, RACHEL TOMPA, PhD, SARAH A. WEBB, PhD, REGINA NUZZO, PhD
- Simulating Crowded Cytoplasm
- Animating Molecular Biology
- Capturing Mitosis Genes in Action
- Cells' Collaborative Middle Management
- Hot Bodies a Lure for Unseen Specks
- Brain Folding

25 UNDER THE HOOD | USING DOMAIN SPECIFIC LANGUAGES TO ACCESS PARALLEL COMPUTING IN ALL ITS FORMS
BY PAT HANRAHAN, PhD

26 SEEING SCIENCE | ARCHITECTURAL COMPUTATION VISUALIZES CELL CHOREOGRAPHY BY KATHARINE MILLER

Cover Art: Created by Rachel Jones of Wink Design Studio using book image, © Arosoft | Dreamstime.com, and DNA Methylation image, © Christoph Bock.





Summer 2010
Volume 6, Issue 3
ISSN 1557-3192

Executive Editor David Paik, PhD

Managing Editor Katharine Miller

Associate Editor Joy Ku, PhD

Science Writers
Katharine Miller, Kristin Sainani, PhD, Louisa Dalton, Rachel Tompa, PhD, Sarah A. Webb, PhD, Regina Nuzzo, PhD

Community Contributors
Atul Butte, PhD, Joy Ku, PhD, Pat Hanrahan, PhD

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 S231
Stanford, CA 94305-5444

Biomedical Computation Review is published quarterly 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)
Jennie Larkin, PhD (NHLBI)
Semahat Demir, PhD (NSF)
Jim Gnadt, PhD (NINDS)
Peter Highnam, PhD (NCRR)
Jerry Li, MD, PhD (NIGMS)
Richard Morris, PhD (NIAID)
Grace Peng, PhD (NIBIB)
Nancy Shinowara , PhD (NCMRR)
David Thomassen, PhD (DOE)
Jane Ye, PhD (NLM)

GuestEditorial

BY ATUL BUTTE, PhD, ASSISTANT PROFESSOR OF PEDIATRICS
AT STANFORD UNIVERSITY SCHOOL OF MEDICINE

Democratizing Integrative Biology

The word Om (or Aum) has many meanings in ancient Hindu philosophy, one of which is "that which contains all other sounds." The meaning has relevance to the now commonly used suffix "-ome", used to describe the nearly-comprehensive cataloging of discrete or countable items from a single vantage point (e.g. genome, proteome, envirome, and others). Incredible discoveries in life science and medicine have certainly come about from the broadening of thinking of translational scientists, from single molecules to nearly-comprehensive sets of molecules, such as the discovery of molecular subtypes of cancers through gene expression microarrays. But there has also been some disappointment, as some aspects of disease remain resistant to understanding through the measurements we intuitively use, like the genetic architecture of complex diseases still hidden from genome-wide association studies. It is for this reason that integration across measurements made from several vantage points may grant us the missing clues towards deciphering still unsolved mysteries in life science and medicine.

Much has already been written about the potential of integrating the results of cross-modality experiments (*1-4*). Vidal and others have noted that integration of multiple functional maps can lead to novel informatics algorithms and findings (*2, 3*). And a number of research groups have integrated some combination of gene expression data, protein interactions, phenotypes, linkage data, and/or RNAi studies to generate results that are relevant to diseases including obesity, cancer, Leigh syndrome, and cardiovascular disease (*5-10*). I mention these examples to illustrate that exploration into the integration of multiple modalities is well underway, and can yield high-impact results even for translational medical research.

However, the challenge is that methods for studying biology in an integrative manner are not yet easily accessible to most clinician scientists interested in discovering disease mechanisms or disease biomarkers. Integrating results across measurement modalities (e.g. RNA and proteins, genotype and RNA, etc.) requires a level of computational sophistication and biological knowledge that is difficult to operationalize today. This lack of tools has its greatest impact on translational research. Though I acknowledge that clinical scientists have a number of other hurdles to overcome in biomedical institutions (e.g., getting research resources and protected time), I believe that deploying web-based integrative biology tools to clinician scientists could enable them to start hypothesis generation and discovery of candidate markers for the conditions they treat.

For example, an interventional cardiologist empowered in this way might be in the best position to ask a novel biomedical question looking for candidate serum markers for coronary artery stent restenosis across diverse biomedical data. Democratizing the process of integrative biology to the clinician scientist, and providing web-based tools operating in the cloud for them to conduct integrative biology experiments using their own data as well as public data, could eliminate one of the remaining bottlenecks in the translational lifecycle.

I encourage computational scientists to consider developing and deploying tools for the quantitative clinician scientist. As another definition of Om is "the essence of the universe," there are still many –omes remaining for translational scientists to explore, integrate, and harness for the improvement of human health.

Democratizing the process of integrative biology to the clinician scientist, and providing web-based tools operating in the cloud for them to conduct integrative biology experiments using their own data as well as public data, could eliminate one of the remaining bottlenecks in the translational lifecycle.

Footnotes

1. R. Paylor, *Nat Methods* **6**, 253 (Apr, 2009).
2. H. Ge, A. J. Walhout, M. Vidal, *Trends Genet* **19**, 551 (Oct, 2003).
3. M. Vidal, *Cell* **104**, 333 (Feb 9, 2001).
4. A. J. Walhout *et al.*, *Curr Biol* **12**, 1952 (Nov 19, 2002).
5. E. E. Schadt, S. A. Monks, S. H. Friend, *Biochem Soc Trans* **31**, 437 (Apr, 2003).
6. E. E. Schadt *et al.*, *Nature* **422**, 297 (Mar 20, 2003).
7. V. K. Mootha *et al.*, *Proc Natl Acad Sci USA* **100**, 605 (Jan 21, 2003).
8. M. Stoll *et al.*, *Science* **294**, 1723 (Nov 23, 2001).
9. D. R. Rhodes *et al.*, *Proc Natl Acad Sci USA* **101**, 9309 (Jun 22, 2004).
10. S. B. English, A. J. Butte, *Bioinformatics* **23**, 2910 (Oct 5, 2007). □

SimbiosNews

BY JOY P. KU, PhD, DIRECTOR OF DISSEMINATION FOR SIMBIOS

Reaching Under the Hood of a 20-year-old Musculoskeletal Model

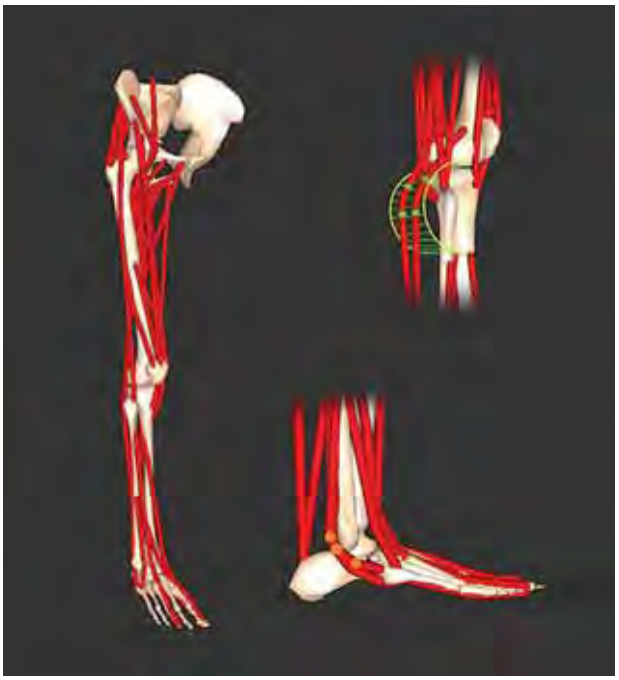
It's often said that all models are wrong, but some are useful. And one model that certainly falls in the "useful" category is the human lower-limb model that Scott Delp published in 1990. It captures the musculoskeletal geometry and muscle arrangements from the hips down to the feet and has been used in numerous simulation studies over the years, providing insights in fields such as sports and medicine. As the saying goes, though, the model is not perfectly accurate. Based on two very small, decades-old studies, the model is limited in how faithfully it can reproduce human movements. Now, with the availability of new experimental data, researchers have updated this lower-limb model, enabling it to address new research problems.

"This is an exciting new tool," says **Edith Arnold**, a mechanical engineering graduate student at Stanford University and the creator of the new model. "It should both correct some of the problems people were having with the old models and allow people to answer new questions."

To gather data for the new model, **Samuel Ward, PT, PhD**, assistant professor in radiology, and his colleagues at the University of California, San Diego took apart 26 human muscles from each of 21 different cadavers, examining them fiber by fiber to determine both their organization and physiological properties. Using a laser technology that was only just emerging when Delp created the original model, the group measured the length of the sarcomeres, the individual subunits that make up the muscle. This critical piece of information allowed them to normalize their muscle fiber measurements and determine the force-length relationship that characterizes how a muscle performs.

The original model is like a car where no one's opened the hood, says Ward. We know the car performs in a certain way but its parts and how they function together have not been well-studied. "We lifted the hood on the model, looked around, and got a whole bunch of new data to really understand the fundamental properties of the model."

Not only is the data more comprehensive, it is also derived from a much larger number of cadavers than was the original model: 21 versus 2 or 3. The larger sample size provides a better idea of what constitutes normal muscle architecture and behavior.



Having a model based on this number of cadavers is great, says **Jonas Rubenson, PhD**, assistant professor in the biomechanics group at the University of Western Australia. "It's a huge leap forward in these models, and now we can be a lot more confident that the muscle parameters are actually representative."

Rubenson is using the model to test some of the assumptions in his experimental study of the force-length properties in the calf muscles. However, Arnold points out that there are a lot of other reasons why people may be interested in the updated model, including a more realistic representation of the knee and changes to bone geometries based on new imaging data.

Creating these models is difficult and time-consuming, says Arnold. "But I like creating these tools that are going to be useful for many people. I'm excited to use this model in my own research and to see what others will do with it." □

DETAILS

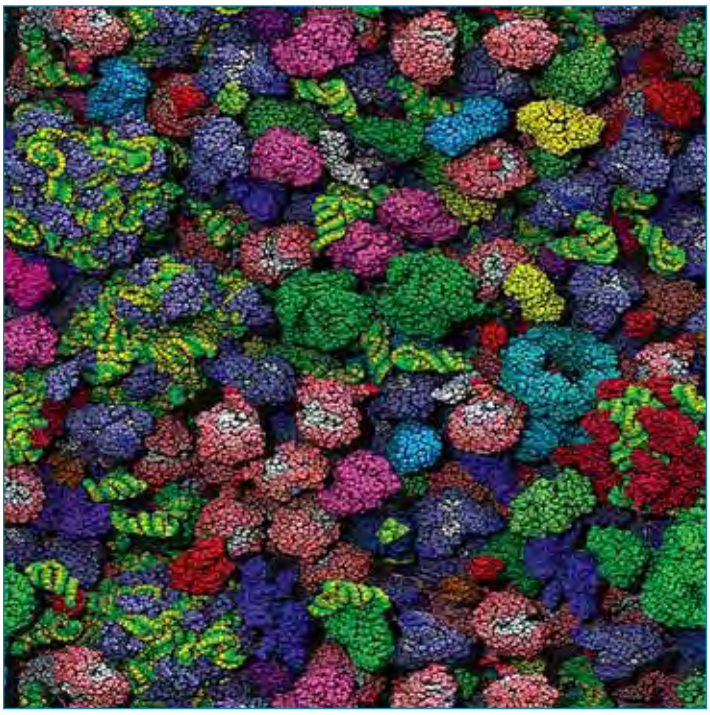
The updated lower limb model can be accessed via the neuromuscular models project at http://simtk.org/home/nmblmodels. This Web site is a simtk.org umbrella project, providing links to many other musculoskeletal models that are available for downloading from the simtk.org website. See also, Arnold, E.M., Ward, S.R., Lieber, R. L., and Delp, S.L., A model of the lower limb for analysis of human movement, *Annals of Biomedical Engineering*, 10.1007/s10439-009-9852-5 (2009).

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

NewsBytes

Simulating Crowded Cytoplasm

In biology textbooks, the carefully rendered cross-section of an *E. coli* cell often resembles a well-organized and spacious apartment, with everything in its place and ample room for movement. But a recent computational recreation of the scene looks more like a Friday night dance floor, with molecules bumped up against one another in every direction. In addition to providing a dramatic, qualitative description of the crowded cytoplasm, this first atomically detailed computational model of *E. coli* innards is also a tool for quantitative predictions of molecular conduct within the cell. The model is described in the March 2010 issue of *PLoS Computational Biology*.



Combining all available known details about the atomic structures and concentrations of 50 of the most common proteins within* E. coli*'s cytoplasm, Elcock and McGuffee created a model of what it might be like inside the crowded cell. They then simulated 20 microseconds of jostling with and without various types of molecular interactions, including crowding (excluded volume effect) and electrostatic and hydrophobic interactions. They then compared the results to experimental observations. Reprinted from McGuffee SR, Elcock AH, 2010 Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm.* PLoS Comput Biol *6(3): e1000694. doi:10.1371/journal.pcbi.1000694.

"This is an attempt to build a virtual lab, in which we can study various biological and biophysical processes as they might occur inside the cell," says **Adrian Elcock, PhD**, coauthor and associate professor of biochemistry at the University of Iowa.

The sea of floating proteins inside every cell is the background against which many cellular reactions take place. Scientists realized years ago that the cytoplasm is generally not an invisible player in those reactions. One of the best-studied examples is macromolecular crowding (also called excluded volume effect). Having large neighbors on every side changes a protein's effective concentration and influences its movement and ability to react. A biological reaction observed in dilute solution can be much faster or slower than the same reaction inside a crowded cell.

To create the model, Elcock and then graduate student **Sean McGuffee, PhD**, started by gathering known structural data for 50 of the most common *E. coli* proteins. They then combined the detailed representations inside a computer model at known cellular concentrations, creating a strikingly dense model of 1008 proteins. The researchers then set that image in motion, running independent Brownian dynamics simulations governed by varying energetic descriptions of intermolecular interactions. The simplest description included only the excluded volume effect: no molecule could take the space of another molecule. The most complex scenario they ran included excluded volume, electrostatic interactions, and favorable short-range hydrophobic interactions. The more complex simulations performed surprisingly well when asked to predict molecular behaviors, such as diffusion and stability, in the *E. coli* cytoplasm.

The model was able to match experimental observations of how quickly green fluorescent protein diffuses in the *E. coli* cytoplasm. And it was able to predict the greater stability of the unfolded state of the protein CRABP, cellular retinoic acid binding protein, over the folded state inside *E. coli*. Although the presence of close neighbors (crowding) generally stabilizes a large folded protein, the specific electrostatic and hydrophobic interactions of unfolded CRABP with other cytoplasmic proteins counteract the crowding effect.

"What this doesn't mean," Elcock emphasizes, "is that crowding effects are unimportant. It means that crowding is only part of the story."

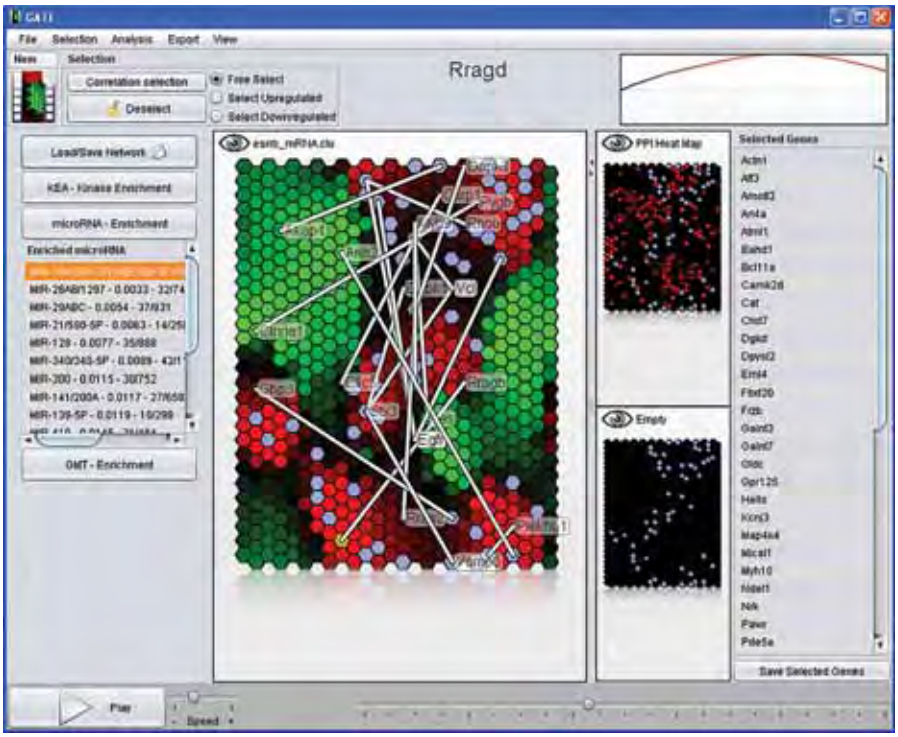
A computational box of 1008 proteins is still a far stretch from the complex *E. coli* cytoplasm, says **Allen Minton, PhD**, a pioneer in the study of crowding effects and researcher of physical biochemistry at the National Institutes of Health. "But there are a lot of questions that only this type of computation can answer," he says. "From a computational point of view, it is a real tour-de-force."

—By Louisa Dalton

Animating Molecular Biology

These days, molecular biologists often gather data over a period of time—observing shifts as they occur inside groups of cells undergoing natural changes. The researchers then face the daunting task of making sense of it all. Now, computational biologists have devised a software program to easily visualize and analyze these mountains of time-series data in animated movie form. While these flicks might never

appear at a theater near you, scientists studying such disparate areas as stem cell development and the microbial communities of the Pacific Ocean will be playing them on their computer screens to explore how all the genes and proteins of a cell type or organism change over the timespan of experiments.



This screenshot from the GATE software program shows RNA expression levels from experiments on stem cells that were genetically manipulated to differentiate. Each hexagon represents a single gene; red hexagons are genes with increased RNA levels and green are those with decreased levels. Commonalities among gene annotations are highlighted in blue, and white lines represent known interactions between proteins. GATE movies animate a series of these images to show changes over time. Courtesy of Avi Ma'ayan.

"This is a tool that is really useful for interrogating datasets collected as a time series at multiple layers of regulation," says **Avi Ma'ayan, PhD**, assistant professor of pharmacology and systems therapeutics at the Mount Sinai School of Medicine who spearheads the project. "It allows you to form hypotheses for future experimentation very quickly."

The software, called GATE (Grid Analysis of Time-Series Expression), was originally designed to analyze clustered gene and protein expression data taken at various time points during stem cell development, Ma'ayan says. This work, led by **Ihor Lemischka, PhD**, Mount Sinai professor of gene and cell medicine, was published in *Nature* in November 2009. "It was a relatively simple approach but it hadn't been done before," Ma'ayan says. But Ma'ayan's group realized that GATE movies would be even more useful if they could incorporate existing biological data, such as libraries of protein-protein interactions or annotations of genes' functions. The updated software was further described in *Bioinformatics* in January 2010.

The movies GATE generates show a 2-D honeycomb of small hexagons, each representing a single gene or protein and colored red (for increased expression) or green (for decreased). The hexagons are clustered near other genes or proteins with similar behavior patterns in the experiments. When the movie plays, waves of color shift across the grid, representing molecular shifts over the time course of the experimental series. Although GATE was developed for stem cell biologists, its potential applications are broad, Ma'ayan says. Recently, he was contacted by a group at the University of British Columbia that wants to use the software to analyze changes in marine flora and fauna in the Pacific Ocean. In this case, the movies will look at changes both over time and distance, as the researchers sample further from the coast.

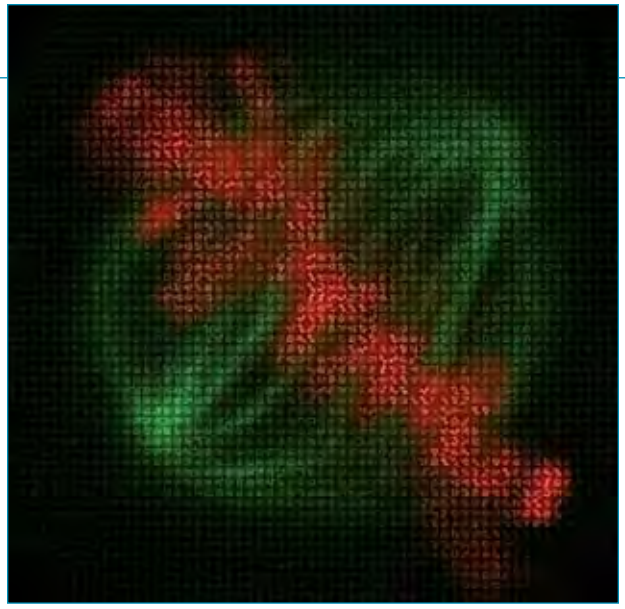
Oliver Hofmann, PhD, a computational biology research scientist at the Harvard School of Public Health, says the technology will be very useful for the field of molecular biology. "It's a very neat way of visualizing time series," he says. "But it's not just a pretty picture you can look at. You can explore it interactively too." It is still difficult to coordinate more than two types of data timecourses in GATE, Hofmann says, and Ma'ayan agrees. He says their to-do list includes plans to better overlay multiple movies.

—By Rachel Tompa, PhD

Capturing Mitosis Genes in Action

During the one-hour drama that is human cell division, many genes enter and exit the stage. Until now, researchers did not know the identities of many of these actors, nor understand their various roles. Now, using a combination of high-throughput screening methods, time-resolved movies and a supervised machine-learning algorithm, researchers have identified 572 genes that are involved in mitosis in human cells. The raw data and images are available to the research community at www.mitocheck.org.

"Researchers can go to the database, do a clustering analysis, and extract the genes that are most interesting from their research question point of view," says **Jan Ellenberg, PhD**, head of the Cell Biology and Biophysics Unit at the European Molecular Biology Laboratory and senior author on the paper pub-



This microscopy image captures the mitotic spindle (green) and the chromosomes (red) of a dividing cell. EMBL researchers videotaped mitosis for 22,000 different gene knockouts. Videos and data for all 22,000 genes are available at www.mitocheck.org. Courtesy of Thomas Walter & Jutta Bulkescher / EMBL.

lished in *Nature* on April 1, 2010.

The research addressed an age-old problem in the study of cell division, Ellenberg says. "We didn't know all the genes or the proteins involved," he says. "So we decided that we had to do this gene discovery ourselves."

First, Ellenberg and his colleagues in the Mitocheck consortium developed the technology to do systematic high throughput screens of multiple samples of all 22,000 human genes and then visually match each knockout to a phenotype. They relied on RNA interference to knock out each of the approximately 22,000 individual genes. They then printed more than 384 of these samples at a time on microarray chips. Because mitosis occurs transiently (approximately once every 24 hours), the researchers developed microscopes to capture movies of each sample from four such microarrays in parallel over the course of 48 hours.

Analysis of so much visual data—nearly 200,000 movies—required supervised machine learning. First, a human expert annotated examples of different morphologies observed within the movies. A computer then extracted a numerical signature with 200 different parameters that it correlated with those characteristics. After iterative training with movies of just 3000 different individual cells, the computer analyzed additional movies and identified phenotypes with 90 percent accuracy. The researchers also developed new distance measures for clustering algorithms to categorize the differences in cell division behavior.

The scale of these experiments and the use of time-lapse imaging over two days are "unparalleled and nothing short of phenomenal," says **Anne Carpenter, PhD**, director of the Imaging Platform at the Broad Institute, who was not involved in the research. "[The insights into mitosis are] just the tip of the iceberg of the knowledge that will be extracted from this single experiment," Carpenter says.

The researchers' next project, called Mitosys, will explore the molecular activity of the 572 mitosis-related genes.

—By Sarah A. Webb, PhD

Cells' Collaborative Middle Management

Like corporate and governmental organizations, cells rely on middle managers to keep things running smoothly. These "middle managers" function as a critical bridge that controls the flow of information traffic. According to recent research, however, the middle managers often partner with one another, ensuring that the failure of one manager doesn't bring down the entire organization. Moreover, this partnering becomes more extensive in more complex organisms.

"Understanding the system isn't about the function of the individual parts. [It's about] understanding the importance of these information flow bottlenecks and how natural systems get around them," says **Mark Gerstein, PhD**, professor of bioinformatics at Yale University. He and his colleagues have been studying networks of genes and transcription factors to describe the information flow within cells.

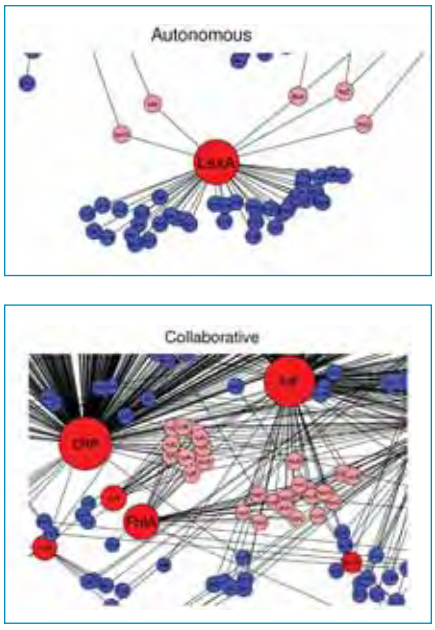
The work serves as part of a larger effort within Gerstein's group to develop real-world analogies to explain how biological systems use and process information. Previously, the group had shown that hierarchies in biological regulatory systems resemble directed social structures such as governments and corporations. That study, published in *PNAS* in 2006, found that "middle managers rule," Gerstein says. Transcription factors in the middle layers of the networks have the most regulatory interactions with other genes. "The genes in the middle are much more essential. If you knock them out, the organism is much more likely to die."

In a paper published online in *PNAS* in March 2010, Gerstein and his colleagues took that work a step farther, seeking to understand how cells avoid failure at the sites of middle manager bottlenecks in five species ranging from *E. coli* to humans. First, they identified which genes are regulated by other genes in each organism. They then stacked the levels of regulators in hierarchies and placed them between two extreme types of social hierarchies, autocratic and democratic, and showed cellular regulatory hierarchies have "intermediate" structures. They found that, in all five organisms, coregulation happens most at the middle level and least at the bottom. And more complex organisms exhibit more collaborative, "democratic" regulatory structures with more interconnections. For example, yeast has about one regulator for every 25 targets whereas in humans the ratio is much smaller, about one to 10.

"The parallels between government structure and regulatory network structure are provocative," says **Trey Ideker, PhD**, associate professor of medicine and bioengineering at the University of California, San Diego, who was not involved with the study. One question, says **M. Madan Babu, PhD**, an investigator in the MRC-Laboratory of Molecular Biology at the University of Cambridge, is the function of these hierarchies within a cell. "Are they really important? Or

are they something that is emergent because of the complexity of the system and has no consequence whatsoever?"

Regulatory networks are definitely important for organism function, Gerstein notes. So the question of whether the networks emerged in response to complex roles or the system's complexity allows organisms to carry on these complex interaction is a "chicken and egg type of issue."



Diagrams of hierarchical networks: In an autocratic network, such as the military, there is a clear chain of command. In a democratic network, many members interact and regulate each other. And in an intermediate network, such as exists within a law firm and many cells, the hierarchy shares features of both types. As biological organisms become increasingly complex, their organization becomes more democratic.

—By Sarah A. Webb, PhD

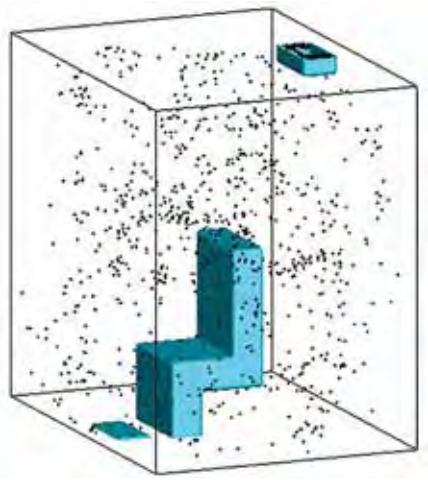
Hot Bodies a Lure for Unseen Specks

We can't see them, but tiny particles—dust, pollen, microbes, and the like—swirl around us in complicated, turbulent pathways. New numerical simulations suggest that, at least in tiny indoor spaces, our body heat may pull them even closer, where they have a better chance of eventually landing in our lungs.

"The conventional wisdom is that the thermal plume from your body protects you from particles falling from above," says **John B. McLaughlin, PhD**, professor of chemical and biomolecular engineering at Clarkson University and coauthor of the study. "We found that, in our small room at least, that is not true." Such findings can help engineers design better ventilation systems, McLaughlin says. "Studies have shown that schoolchildren learn more and office workers are more productive in environments where the concentration of particles in the air is very low."

Airflow dynamics are notoriously tough to model computationally, largely because of the huge range of physical scales in equations for turbulent fluids. McLaughlin and his colleagues used a direct numerical simulation approach that offers accuracy but requires intensive computational resources. Their computational models of airflow and particle paths were built in a 4.8-square-meter virtual room at two-centimeter resolution using three-millisecond time steps over about three minutes of total simulated time. In each simulation, a mannequin sits motionless in the middle of the room. A stream of air suffused with particles—each with the density of sand and about the size of a grain of pollen—shoots up through a floor vent in front of the chair. Particles fan out throughout the room, with a ceiling vent as the only exit.

In simulations where the mannequin was bestowed with realistic body heat, researchers could see the hot air surging off the body and interacting with particulates. This thermal plume pulled rising particles directly into the mannequin's breathing zone. At the same time, the plume blocked the path of particles traveling near the ceiling, forcing them to fall down into the mannequin's personal space, doubling the trapping effect of the plume. The work was presented in March 2010 at the American Physical Society meeting in Portland, Oregon.



The positions of 2-micrometer particles inside a 20-degree-Celsius room with a mannequin heated to 25 degrees Celsius, three minutes after particles were released through a floor event. In this simulation, 31 out of 1000 particles fell directly onto the mannequin's warm body; none managed to leave the room through the ceiling vent. Yet when the mannequin was the same temperature as the room, no particles fell onto the body, and 160 out of 1000 particles escaped. Results were similar for simulations with 10-micrometer-diameter particles.

"The computational and the experimental go hand in hand when studying complex turbulent flows such as those around human beings," says **Mark N. Glauser, PhD**, professor of mechanical and aerospace engineering at Syracuse University, whose empirical results helped guide McLaughlin's modeling. Fundamentally, experiments can help validate computational models and give physical insights that spur new simulations. "Then the simulation tools can be used to probe a broader range of parameter space 'virtually,' as well as look in more detail at flow physics," Glauser says. For example, the models from McLaughlin's team can track individual particles in a turbulent flow—a feat that's nearly impossible in real-life experiments.

—By Regina Nuzzo, PhD

Brain Folding

In the four months before birth, a fetus's brain grows from a smooth tube of neurons into a highly crinkled, convolved mass of tissue. Because the cerebral cortex has a surface area nearly three times as big as that of its skull cavity, scientists have proposed that this real-estate-space squeeze is what drives the brain's folding process. Now results

from a computational three-dimensional geometric model agree that the skull does help guide the wrinkling—but they also suggest that a growing brain folds up regardless of its container.

"Mechanical constraints imposed by the skull are important regulators," says **Tianming Liu**, **PhD**, assistant professor of computer science at University of Georgia and lead author on the study, which was published in May 2010 in the *Journal of Theoretical Biology*. "But our simulations indicate that skull constraint is not necessarily the dominant mechanism."

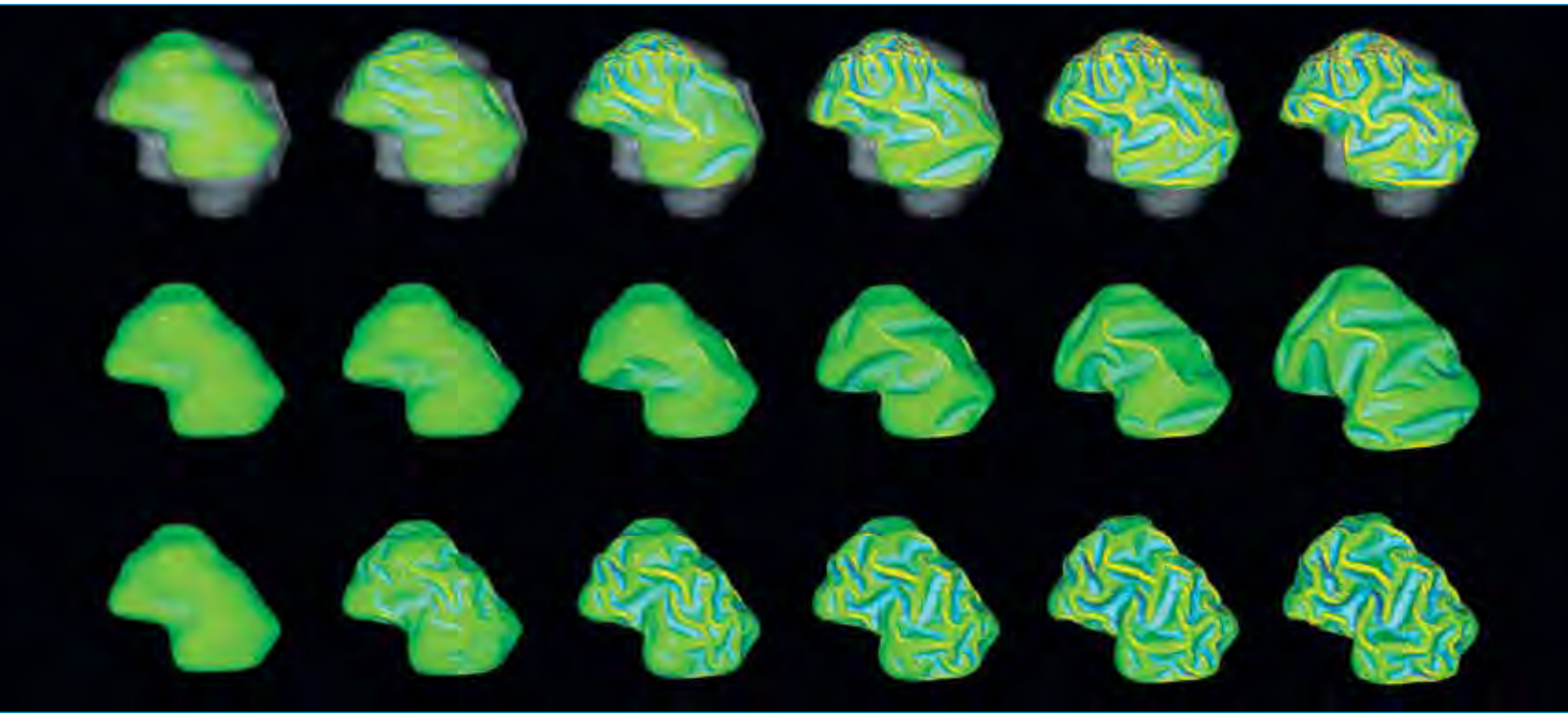
The computational model underneath the simulations had two main features: geometric constraints of the skull, and partial differential equations that model biological processes driving the growth of neurons. To start off the simulation, researchers used MRI data from the brains of two human fetuses; then solutions to the differential equations guided the changing surface geometry of the cortex.

The team simulated how the cortex grows under various conditions: without a skull, with a skull of fixed size, and with a skull that grows at the same time as the brain does. As expected, brains grown in a skull were more convoluted than those allowed to develop unfettered. But even without a skull to confine it, a cortex will still fold in on itself, results showed. This happens as a natural response in a fast-growing cortex, as the tissue attempts to reduce the increasing mechanical tension among axons, dendrites, and neuroglia, Liu says.

Tweaking other parameters in the model revealed how cellular growth affects these folding patterns. When neurons themselves grow rapidly — during synapse development and neuron dendritic projection, for example— the cortical folding increases dramatically too. And when certain areas of the cortex grow more quickly than others, this imbalance subtly shapes what kinds of folds become most prominent.

Computational models can help explain normal brain development as well as what happens when things go wrong, says **Bernard S. Chang**, **MD**, assistant professor of neurology at Harvard Medical School. For example, in some forms of microcephaly, the brain surface is almost completely smooth with no folds; in others, the folding is normal. "A model that predicts how folding is affected by the skull's physical constraints might help us to understand why some patients have one form and not another," he says. Since animal models don't capture the complexity of the human brain, and doing repeated MRIs of developing fetuses for research isn't feasible, Chang says, "we need to rely on these theoretical models as tools to help us understand what we're observing clinically."

—By Regina Nuzzo, PhD □

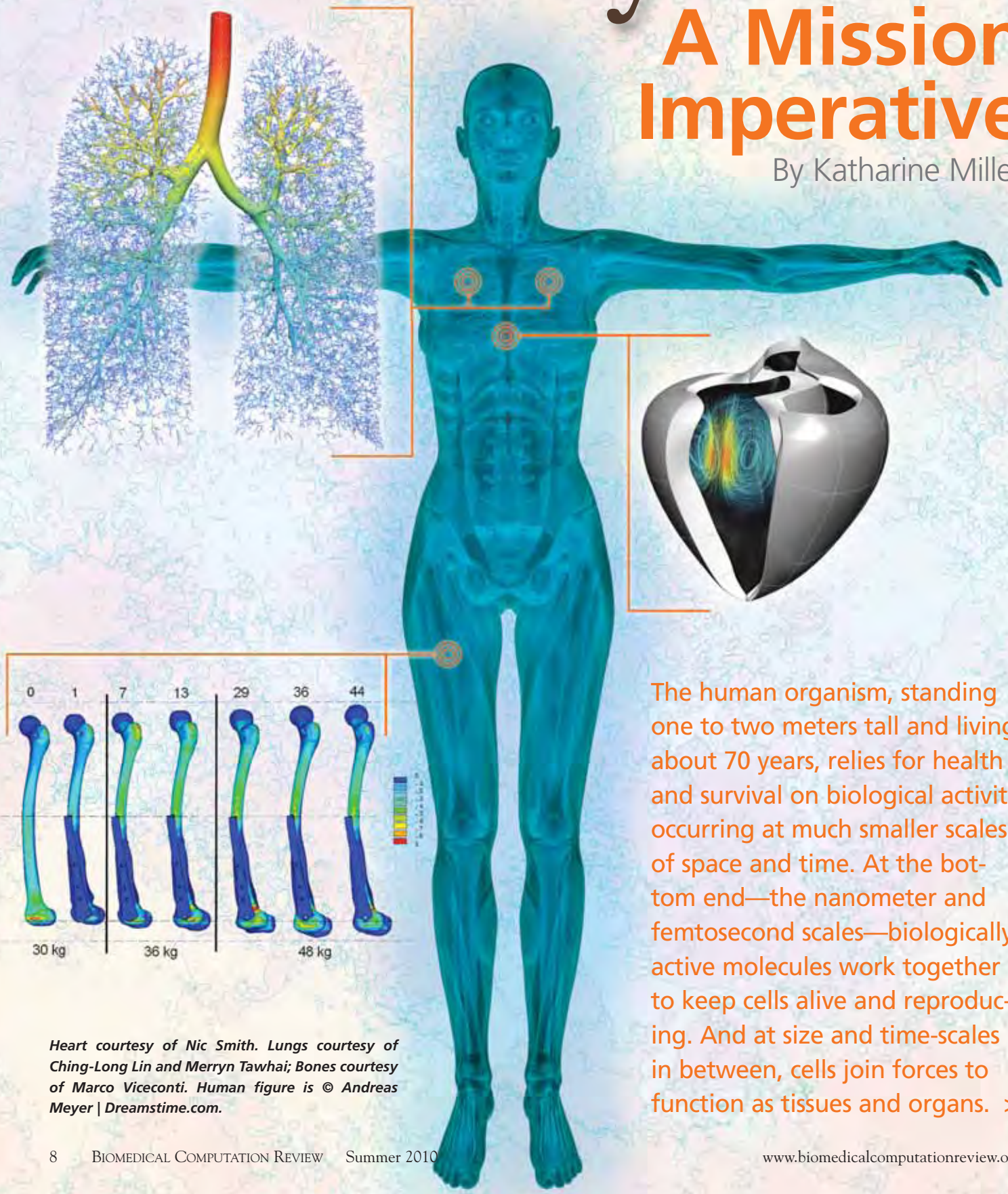


Growth of the cortex under different assumptions. From left to right, the images show simulated development of the cortex over time. The cortex grows (a) within a skull of fixed size, (b) without a skull, or (c) within a skull that also grows at the same time (which corresponds to a fetus' developing skull). Major cortical folds developed much earlier and faster during simulations with skull constraints. But the cortex increases its surface area and convolutes itself to reduce the fast-growing internal tension, with or without skull constraint.

THE Physiome

A Mission Imperative

By Katharine Miller



Heart courtesy of Nic Smith. Lungs courtesy of Ching-Long Lin and Merryn Tawhai; Bones courtesy of Marco Viceconti. Human figure is © Andreas Meyer | Dreamstime.com.

The human organism, standing one to two meters tall and living about 70 years, relies for health and survival on biological activity occurring at much smaller scales of space and time. At the bottom end—the nanometer and femtosecond scales—biologically active molecules work together to keep cells alive and reproducing. And at size and time-scales in between, cells join forces to function as tissues and organs. >

This is the reality of human biology: events span a 10^9 range in lengthscale (molecular to organismal) and a 10^{14} range in timescale (molecular movement to years). To understand this biology—and provide appropriate medical care—scientists need to understand the interactions across these scales.

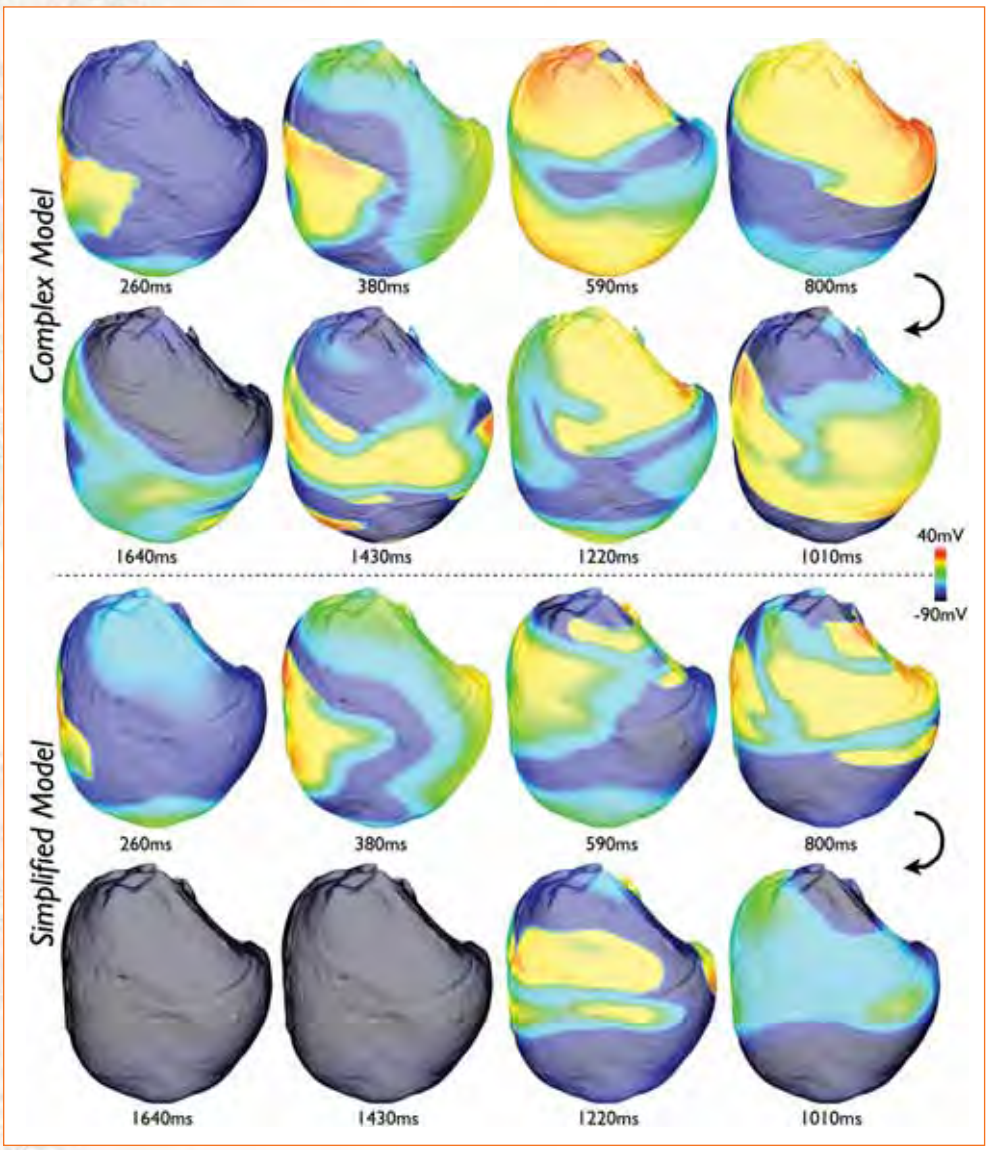
"Systems that have clinical relevance and involve how to treat or prevent disease are always multi-scale and multi-feedback," says **Peter Kohl, MD, PhD**, reader in cardiac physiology at the University of Oxford and a coordinator of the Virtual Physiological Human Network of Excellence (VPH NOE) funded by the European Commission.

Hence the physiome: an international effort to quantitatively describe human physiology across this vast range of scales. "Basically, with the range of scales involved, you have a mission impossible in front of you," Kohl says. "On the other hand, whether you get your head around it with or without quantitative models, it remains the same range. It's therefore a mission imperative rather than a mission impossible."

Fortunately, the goal is not to simulate an entire physiological human on a computer in full detail (which would require more computational power than is available on the planet), but rather to develop models that "simplify the system to provide insight and identify causal relations," Kohl says.

Some physiome models are already providing remarkable insight, says **Jim Bassingthwaighte, PhD**, professor of bioengineering at the University of Washington, who defined and named physiome in the early 1990s. Physiome efforts have sprung up in many different countries, with projects involving just about every organ in the body. Models of the heart are the most advanced, and are currently being used in clinical studies to optimize treatments for a variety of heart problems. Physiome models of the lung and neuromuscular system are also making breakthroughs. **Peter Hunter, PhD**, director of the Auckland Bioengineering Institute (ABI) at the University of Auckland in New Zealand., believes that, over the next few years, multi-scale approaches will achieve clinical outcomes that wouldn't have been otherwise possible.

One reason the physiome can make a difference is that medical treatments are often built on a phenomenology of "I do this thing and get this result," says **Nic Smith, PhD**, professor of computational physiology at Oxford University and scientific coordinator of euHeart, the VPH heart physiome project. "In many ways, the physiome effort is trying to change that: To underpin the result with mechanisms that we can identify and think we understand."



A WORLDWIDE EFFORT

The notion of the physiome, Kohl says, is "more a philosophy than a project, and there are many people around the world who have adopted that philosophy." The European Commission funds the "Virtual Physiological Human" (VPH) effort; the National Institutes of Health in the United States have the Interagency Modeling and

When complex (top two rows) and simple (bottom two rows) rabbit ventricular models were each induced into arrhythmia, the patterns of electrical activity differed noticeably. The complex model included finer anatomical features such as vessels and endocardial structures. These extended the period during which arrhythmic activity was observed (compare last two panels of each sequence), illustrating the importance of using realistic heart models for patient-specific diagnosis and treatment prediction. Simulations performed with the Cardiac Arrhythmia Research Package. Figure courtesy of Dr Martin Bishop, University of Oxford. For further detail see Bishop M., et al., Development of an anatomically detailed MRI-derived rabbit ventricular model and assessment of its impact on simulations of electrophysiological function, **Am J Physiol Heart Circ Physiol *298:699-718, 2010.***

Analysis Group and the Multiscale Modeling Consortium with specific funding initiatives directed at physiome research. The International Union of Physiological Scientists houses "The Physiome Project" and in particular the CellML project (www.cellml.org), fed largely by Peter Hunter's Institute in New Zealand. And countries such as the United Kingdom, Japan and China also have their own physiome projects (not described further here, but meritorious in their own right).

Though they share a similar philosophy, the various efforts have quite different focuses. For example, Peter Hunter's project in New Zealand seeks to build models of every organ system and every level within each organ system. In the future, the plan is to have modules that can be shared and connected to study whatever someone wants to study.

Meanwhile, Europe's VPH project, formerly called the Europhysiome, puts a strong emphasis on engaging industry and clinical centers, says Smith, because that's what interested the European Commission. "It's quite translational and outcome focused, which has really moved the physiome from what is a very appealing scientific vision to being something that might really matter to people," he says. "I think that has been a very positive thing."

In the US, the focus has been more on basic science, says **Grace Che-Yaw Peng, PhD**, program director at the National Institute of Biomedical Imaging and Bioengineering at the National Institutes of Health. "US investigators are digging deeper within each scale or between scales but not necessarily reaching that far out to clinic or industry," she says.

These complementary approaches may help ensure the best research outcomes, Peng says. "Should researchers in Japan, China, US, and Europe agree on the heart model that gets incorporated into clinical process?" That's not necessarily the goal, she says. "Everyone has a different question."

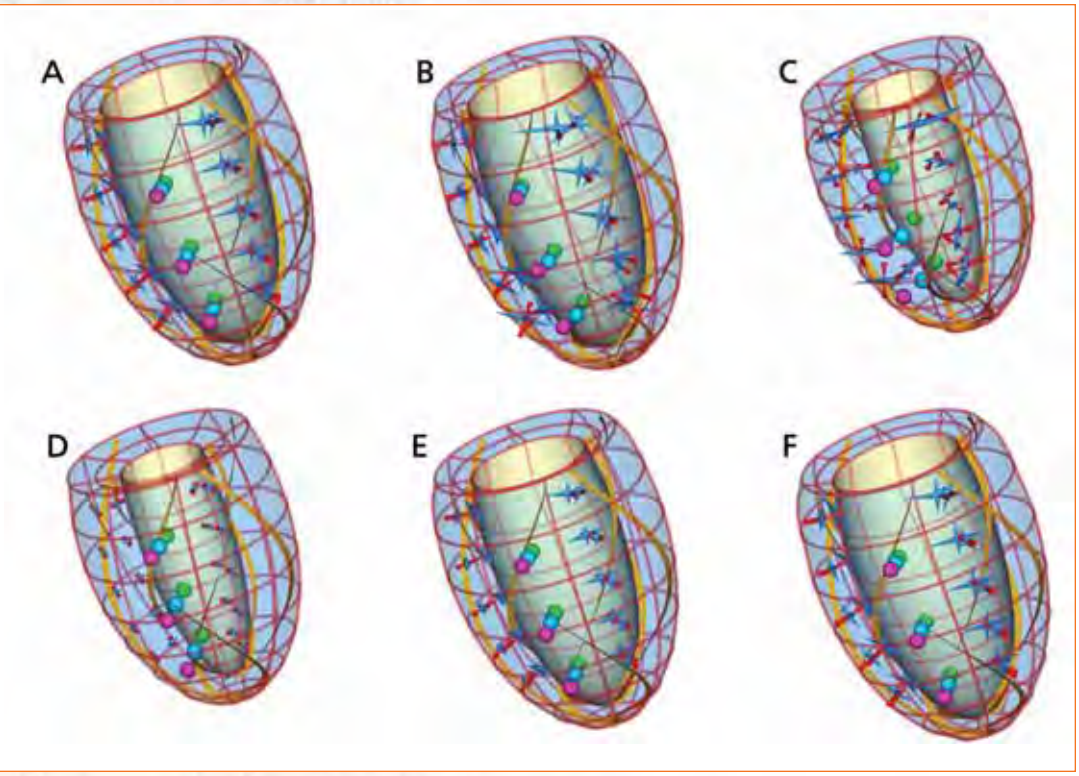
Though a more coordinated approach might be more efficient, there is a precedent for grassroots efforts producing a valuable result: the human genome project. "It wasn't coordinated at the outset, and there was no prescribed effort to control overlap," Kohl says, "yet it's been one of the most successful integration activities worldwide, sharing information and data. The physiome effort can learn a lot from that approach."

"[The Virtual Physiological Human project] is quite translational and outcome focused, which has really moved the physiome from what is a very appealing scientific vision to being something that might really matter to people," says Nic Smith.

THE HEART PHYSIOME

The heart physiome project began 20 years ago as a collaborative endeavour between Auckland and Oxford. "In the physiome vision," Smith says, "the heart is arguably the most advanced example of taking information from lots of sources and putting it into a consistent framework that you can probe in ways that you can't think about all at once."

Although the heart is in some ways simple—it's a pump—it is nevertheless complex as it depends on electrical activation, mechanical contraction, and fluid dynamics. This is perhaps why multi-scale research related to the

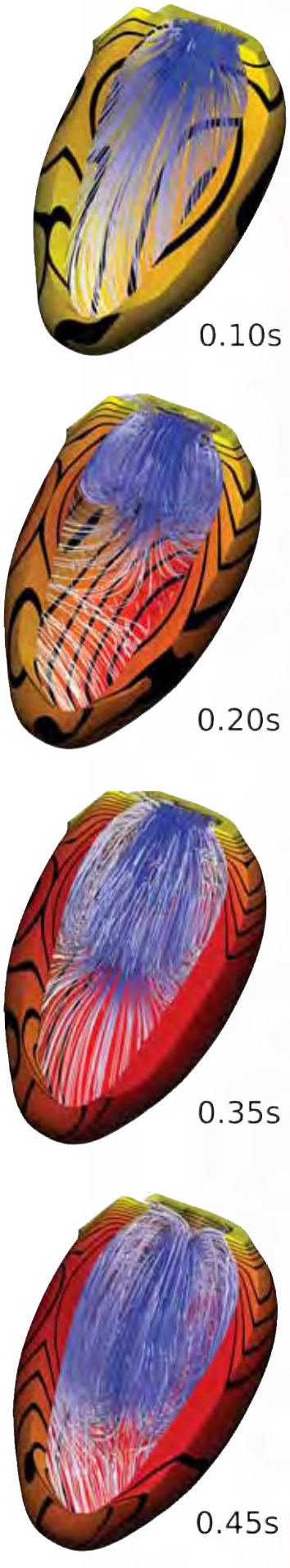


Here, a multi-scale biophysical electromechanics model of the rat left ventricle progresses through a single heartbeat cycle from the end of diastole (A) through multiple key steps including (B) end iso-volumetric contraction, (C) end ejection, (D) end relaxation, (E) end recoil and (F) end diastases. The orientation and size of the cones embedded within the mesh indicate the direction and magnitude of principal strain, respectively. Blue and red cones represent states of compression and tension, respectively. Gold streamlines indicate the fiber orientation. The 3 colored spheres assist in visualizing the rotation of the ellipsoid. Researchers used this model to investigate how feedback loops regulate heart contraction. Reprinted from: Niederer SA, Smith NP, 2009, The Role of the Frank–Starling Law in the Transduction of Cellular Work to Whole Organ Pump Function: A Computational Modeling Analysis.* PLoS Comput Biol *5(4): e1000371. doi:10.1371/journal.pcbi.1000371.

heart has moved farther and faster than it has for other organs. "Lots of different people can all offer a piece of the puzzle," Smith says, and none can understand it alone.

Heart physiome models integrate an impressive number of scales and data types. For example, to understand how a mutation in the myosin regulatory light chain filters up through the scales to alter the dynamics of the heart beat, producing heart failure in humans and in a genetically engineered mice model, **Andrew McCulloch**, **PhD**, professor of bioengineering at the University of California, San Diego and his colleagues created a multi-scale computer model of the mouse heart. The ingredients included models of molecular motors, whole cell twitch forces, tissue stresses and 3-D muscle fiber stresses, ventriclular geometry, and hemodynamic loading conditions. The output showed that changes in phosphorylation of the regulatory light chain (due to a mutation) reduce the twist of the mouse ventricle during systole, which can be an early indicator of heart failure. The changes were then validated in mice *in vivo*.

The heart provides a great testbed for application of the physiome approach to direct outcomes, whether clinical or commercial. For example, if the heart stops contracting synchronously and loses efficiency, patients may benefit from a treatment called cardiac resynchronization therapy: an implanted pacemaker is used to help both ventricles of the heart contract simultaneously. But only about two-thirds of patients given resynchronization therapy respond. So McCulloch's team is developing heart models to optimize this therapy for individual patients. The models rely on patient-specific data that is clinically availible: catheter systems map the heart's electroanatomical activity and hemodynamics; echocardiography measures heart function; and CT scans measure heart structures. The resulting model reconstructs the heart's baseline function and predicts the likely outcome of specific cardiac resyncronization therapy plans. These predictions are then compared to the patient's outcomes at three months. "At this point, we are just following diagnosis and treatment to see if the model can predict what actually occurs," McCulloch says. "If we could better identify responders, or could better identify ways to optimize the therapy to improve results, then the models could have clinical predictive utility."



EuHeart, the VPH program for advancing the heart physiome, is also developing physiome models for optimizing cardiac resynchronization therapy in particular patients, Smith says. "The clinical decisions are in many cases still relatively high level. For example, 'should we put the pacemaker at the back or the front,'" Smith notes. "This means there are often big clinical windows. We don't have to get it exactly right straight away. What we do need is to demonstrate an improvement over best practice. I think this is now possible in an increasing number of contexts."

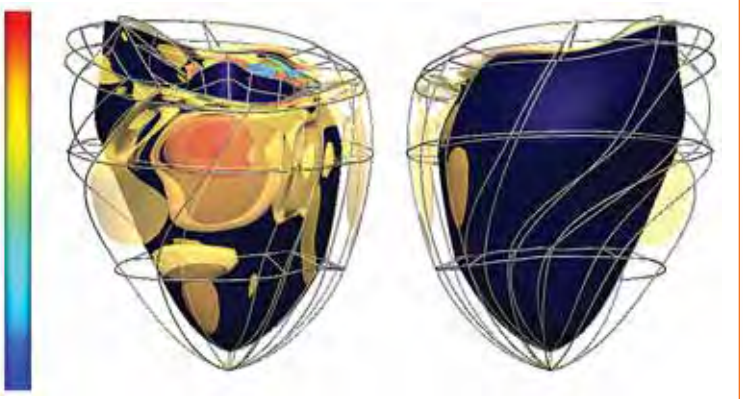
"One of the huge problems with clinical studies is patient variability," McCulloch says. "Any one model is not going to be that useful. We need to understand sources of variation: What's different such that some patients respond and some don't?"

McCulloch and euHeart researchers are now working together to share clinical data. "One of the huge problems with clinical studies is patient variability," McCulloch says. "Any one model

When a large deformation mechanical model of the heart is coupled to a Navier-Stokes model of blood flow within its chambers, researchers can track streamlines through the ventricular volume and observe deformation of the finite element geometry. Courtesy of Nic Smith, Matthew McCormick and David Nordsletten at Oxford University in the UK.

is not going to be that useful. We need to understand sources of variation: What's different such that some patients respond and some don't?"

Another euHeart project models patient-specific coronary blood flow to help doctors determine whether the best treatment for a blockage is a stent, medical therapy, or angioplasty. "We want to be sure we've chosen the right therapy for the right patients," Smith says. Through euHeart, Smith has funding to simulate a number of different patient cases. "Our goal is to get to the point where we have compelling evidence to do clinical trials," he says.



At mid-contraction during an ischemic event, this model of coronary perfusion within the cardiac ventricles shows large gradients in the concentration of oxygenated blood delivered to the heart around the ischemic region but relatively constant perfusion elsewhere. Courtesy of Nic Smith.

Other members of euHeart are determining how to best stop atrial fibrillation. If medication is not successful, the typical treatment is ablation—essentially burning scars in the heart muscle to block the wave of electrical conduction so that it doesn't end up in a cardiac spiral. Currently, more than 25 percent of treated patients have to come back for additional treatments after about three months, says **Olivier Ecabert, PhD**, of Philips Research Laboratories in Aachen, Germany, who is also joint coordinator of euHeart with Smith. The problem is: where to burn? "Ideally," Ecabert says, "doctors would have a patient-specific model and simulate several ablation line options to see how the patient will recover." The physician could then select the ablation line that seems most promising or, pushing it further, the computer could optimize the surgical plan.

Other euHeart projects include predicting when valves are wearing out and should be replaced; and figuring out how to make left-ventricular assist devices (LVADs) work best for the patients who have them.

Philips Research Laboratories, euHeart's project coordinator and one of its industrial partners, joined euHeart because the models might result in some kind of "proof of principle" for software or hardware that Philips could then develop. For example, Philips already developed a geometric heart model that can be adapted to automatically analyze data from 3-D images of the heart. "Now we would like to learn what is necessary to integrate physiological information into the model and then incorporate that into imaging equipment software," Ecabert says.

The fact that industry is interested suggests that physiome modeling, at least of the heart, is coming of age. "We are convinced that this type of physiological model will be more and more applied in the future by clinicians. It's on the rising side of the curve and Philips would like to join the trend early," Ecabert says.

"We are convinced that this type of physiological model will be more and more applied in the future by clinicians. It's on the rising side of the curve and Philips would like to join the trend early," Ecabert says.

THE LUNG PHYSIOME

The lungs sit within the chest cavity where they expand and recoil as the diaphragm contracts and relaxes. Embedded within the sponge-like tissue of the lungs is a branching tree of conducting airways that expand and recoil with each breath. Air flows in through the larger to smaller branches to reach the gas exchange surface, and back out again. Gas exchange—which occurs in the alveoli—also requires a matching supply of blood to and from the gas exchange surfaces.

Multi-scale models of the lung's complexity are starting to yield some interesting findings but are still a step or two away from clinical application. One group of collaborating researchers in New Zealand and Iowa are coupling subject-specific imaging with geometric lung modeling and computational fluid dynamics. The result is a multi-scale lung model that incorporates the entire airway from the oropharynx to the alveoli.

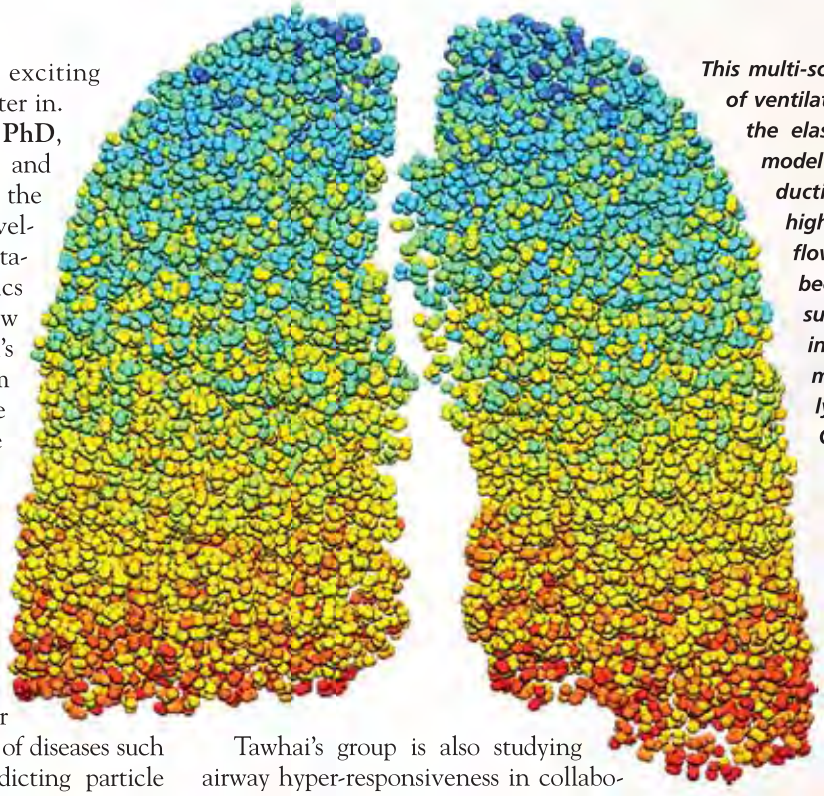
"We're really taking a very systematic and structured approach, similar to what's been done with the heart, in creating anatomically realistic models," says **Merryn Tawhai, PhD**, associate professor in the Auckland Bioengineering Institute at the University of Auckland, New Zealand. "We're building up patient-specific databases and then working down toward modeling cellular functions and putting that into our whole organ model."

One of Tawhai's collaborators is **Eric Hoffman, PhD**, professor of radiology, medicine and biomedical engineering at the University of Iowa. He acquires detailed images of the lung using state-of-the-art spiral computed tomography imaging. The imaging is then converted into a 3-D mesh model of the airway down to the 28th generation of branching using a combination of imaging and mathematical algorithms. This provides a far more realistic image of the airway tree than previous lung models, which typically extend to only the 6th-9th generation at most. The uppermost airways in the model are shaped specifically to match the individual subject's imaging. To fill in the remaining lung tissue down to the airways just before the alveoli, the team uses a volume-filling approach developed by Tawhai's lab and previously validated. Next comes the computational fluid dynamics, to look at airflow, and regional ventilation. And

this is where some exciting results are starting to filter in.

Ching-Long Lin, PhD, professor of mechanical and industrial engineering at the University of Iowa, developed a parallel computational fluid dynamics model to predict airflow in Hoffman and Tawhai's model lung. The team demonstrated that the model can capture detailed flow structures in regions of interest, and can match experimental studies of regional ventilation for the whole lung in a subject-specific way.

The complete model should prove useful for studying the progression of diseases such as asthma and for predicting particle deposition in individual patients, which is important for dosing of inhaled medication, Lin says. One problem, however, is computational cost. To capture laminar-transitional-turbulent flow in the multi-scale airway model required about one week on the TeraGrid Lonestar and Ranger clusters at the Texas Advanced Computing Center for one human subject. "In terms of clinical applications we have to reduce that," Lin says. "Doctors probably don't want that much detail anyway. But the multi-scale computational framework of the human lungs can provide the detailed information needed to understand the interplays between pulmonary structure and function at their most fundamental level."



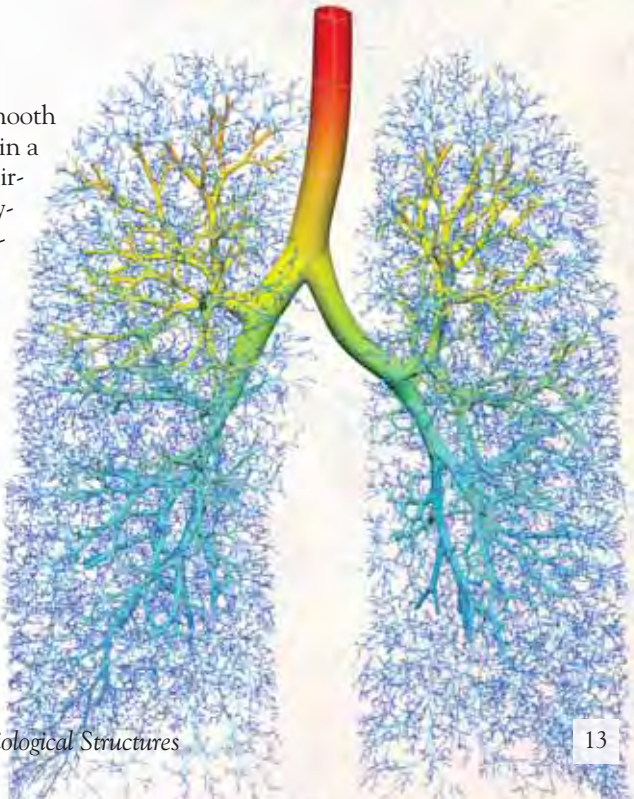
This multi-scale model displays the distribution of ventilation in the lung. The model couples the elasticity of the alveolar tissue to a model of airflow through the entire conducting airway tree. Red represents highest flow; blue represents lowest flow. The vertical distribution occurs because deformation of the lung tissue under its weight makes the tissue in the base of the lung effectively more compliant, so it expands readily when the lung breathes in. Courtesy of Merryn Tawhai.

Tawhai's group is also studying airway hyper-responsiveness in collaboration with several other groups. Together with **James Sneyd, PhD**, professor of mathematics at the University of Auckland, they've developed a model of contraction within the airway smooth muscle cell that is then embedded in a model of the cross section of the airway and the surrounding parenchymal tissue. This is in turn embedded within the whole anatomically structured airway tree model, which is embedded in the lung tissue. "It produces a lung that breathes and develops different forces depending where you are within the lung and so each airway experiences its own particular force balance," Tawhai says.

Experiments by collaborators at the University of Massachusetts, McGill University, and the University of Vermont informed and validated the model. "So we're starting to get a handle on the emergence of patterns of broncho-constriction within the lung and how those vary in different parts of the lung," she says. "Some parts are more susceptible to airway closure than others. This is ongoing work."

In a different project, Tawhai's lab is trying to understand the safest level of heat and humidity for air delivery to mechanically ventilated patients (when an endotracheal tube bypasses the nose and mouth). To get at that question, they had to start at the cell level inside the lungs.

"[T]he multi-scale computational framework of the human lungs can provide the detailed information needed to understand the interplays between pulmonary structure and function at their most fundamental level," says Ching-Long Lin.



A multi-scale, anatomically based mesh of the airway tree. Courtesy of Ching-Long Lin and Merryn Tawhai. Reprinted with permission from IEEE, from Lin, C et al., Computational fluid dynamics,* Engineering in Medicine and Biology Magazine, IEEE, *Issue 3 (25-33) May-June 2009.

On top of the ciliated epithelial cells that line the airway, there's a layer of liquid that must be maintained to a very specific depth in order to achieve mucus clearance. **Nicolas Warren, PhD**, a graduate student in Tawhai's group in the ABI and co-supervised with **Edmund Crampin, DPhil**, from the ABI , developed and validated a model of such cells joined together with liquid moving through multiple cells. Tawhai's team then put the cell model into the whole organ model, distributing cells along airways and through the airway tree, and then directing the lung to breathe with different temperatures and humidity. They found that the epithelial cells alone couldn't transport enough moisture to maintain the depth of the surface liquid during normal breathing. "So there has to be some other significant source of moisture," Tawhai says. "And it's something we couldn't have seen without putting it into the real anatomical framework." Possibly submucosal glands or transport of fluid from the lung periphery provide the additional fluid needed, Tawhai says, but it's really not known. Still, now there's a model on which experimentalists can test various hypotheses. Tawhai's team is currently working on adapting the epithelial cell model to make it more specific to disorders such as cystic fibrosis.

The epithelial cell model is also the starting point for a new NIH grant led by Lin. It will integrate *in vitro* cell data and *in vivo* image data together with Lin's in-house computational fluid-structure-interaction technologies and the cell model to understand the interplay between organ, tissue, and cells. A predictive computational lung model across these scales will allow researchers to assess individuals' response to therapy over time. Ultimately, Lin says, "We will be able to use this information to better tailor a treatment plan for the individual at the most basic level."

THE MUSCULOSKELETAL PHYSIOME

Physiome efforts for neuromuscular modeling are ramping up. A relatively new and major effort is Europe's VPH Osteoporosis project (VPHOP), a collaboration among 19 European academic and industrial partners, led by **Marco Viceconti, PhD**, technical director of the Medical Technology Laboratory at the Istituto Ortopedico Rizzoli di Bologna in Italy. The project seeks to predict the risk of fracture in people with osteoporosis. As people age, their bones weaken and lose calcium, causing a condition known as osteoporosis. Meanwhile, they lose neuromuscular control, which can lead to falls. These changes happen at the cellular level in the bones and muscles and manifest as changes in morphology at the tissue level. "So in order to predict risk of fracture over time, you have to account for whole body, organ and tissue scales," Viceconti says. "That's what we're doing in VPHOP."

By September of this year, two years into the VPHOP project, Viceconti expects to run a very large probabilistic model that accounts directly or indirectly for all factors that act or contribute to the risk of fracture in one patient at any possible scale. The simulation should answer the question: "Of the dozens of possible parameters that can define the multi-scale phenomenon, which ones really are important and make a difference?" he says. "That answer will drive the most critical part of the project—not the modeling itself, but the ability to measure in patients the information we need with the accuracy we need."

The VPHOP is partnering closely with industry to develop the technologies for measuring this key patient information in a cost-

Standardizing the Physiome

Multi-scale quantitative models need to be validated and reproducible if they are to be useful for clinical workflows, says Hunter. The Physiome infrastructure developed by Hunter, Dr Poul Nielsen and their colleagues (and provided at www.cellml.org) makes that process more robust and transparent, he says. Researchers can confidently download an annotated model from www.cellml.org knowing that it's reproducible. The model can then be incorporated into larger scale workflows for use in a clinical setting.

"Having the means to incorporate the outputs of different groups through standards and interoperability is quite a worthwhile goal," Hunter says. "And an essential one if we're to get the modeling of biology into clinical processes."

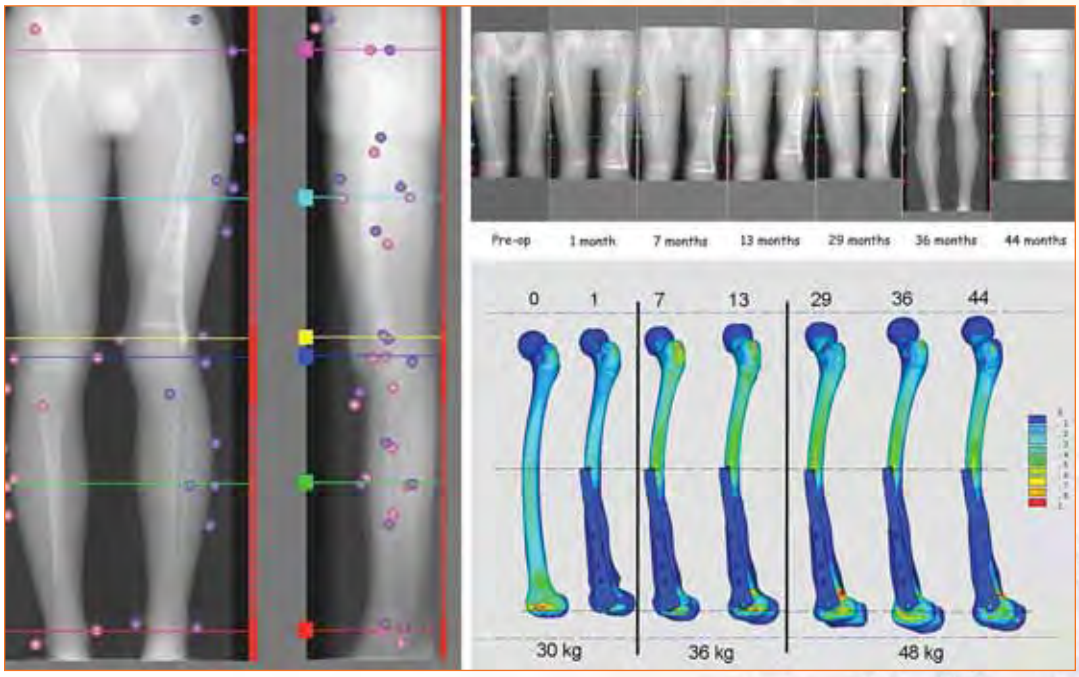
Models held by the models.cellml.org model repository use CellML, a markup language for biophysical models of cells. A repository at the European Bioinformatics Institute (EBI) contains models marked up with SBML, a language for systems biology models. Hunter's group is also creating a new standard called FieldML for integrating spatial information. In recent years, Hunter says, the CellML and SBML communities have become more integrated. "SBML and CellML are now working together jointly to curate models and develop standards around metadata."

From funding agencies' point of view, "We don't want people to have to reinvent models," Peng says. But at this point, "The different formats are all co-existing. No one wants to stand up and say one is better than another."

It's also true that some multi-scale models require information that goes beyond what CellML or SBML can provide, McCulloch says. "It's not possible to describe everything in our cardiovascular model using that system." So McCulloch is building a database that includes metadata about his models that will be consistent with CellML and other model description formats but goes beyond them to include additional information.

Nic Smith agrees that standards are useful for sharing between different academic centers, but he says, an important step to embedding multi-scale models in clinical workflows is a demonstration that they add extra information that can be made available to physicians in a familiar format. "We are working on developing interfaces and putting them in a context where physicians are used to seeing them—in connection with imaging and clinical data accessed directly from the hospital's computer system."

effective way. For example, they've developed ACTIBELT, a device embedded in a belt buckle that can record the kinematics of the body for five days. Also, jointly with Philips, VPHOP is developing a system, based on Philips Medical Systems XpertCT imaging technology, that can generate 3-D images of bone at the tissue scale—primarily for patients at high risk for a fracture. And with BioSpaceMed, they are building a very low-dose whole body X-ray machine called EOS-QT that can generate a 3-D model of the patient skeleton using sophisticated morphing technology and possibly even estimate densitometry at each point and provide a preliminary estimate of the risk of fracture. "We are trying to push limits of the current imaging technology by using all possible tricks," Viceconti says.



A clinical application of a multi-scale model is used several times a year at Viceconti's Institute to help monitor children who've received bone transplants as a treatment for a rare type of bone cancer called Ewing's sarcoma. These children need aggressive rehabilitation, but doctors don't want to risk fracturing the reconstructed bone. So Rizzoli bioengineers do a full scan and gait-analysis with markers, collecting all possible data over a whole day. They then generate a whole body and organ level model for the bone and simulate rehabilitation exercise to predict loads acting on the bone and determine fracture risk. "With this information, we can assist in determining the rehabilitation program for each patient," Viceconti says. "This is the only real-world application we have in clinical practice today." Courtesy of Marco Viceconti.

The VPHOP project has recently started a cooperation project with Simbios, a National Center for Biomedical Computing at Stanford (and publishers of this magazine). Much of Simbios' neuromuscular modeling work has a multi-scale aspect, which opens the door to musculoskeletal physiome modeling. VPHOP and Simbios hope to connect their online communities and integrate their tools. Eventually, Viceconti says, "we'd like to join forces to attack a grand problem where the multi-scale approach can make a difference."

For example, both Viceconti and **Scott Delp, PhD**, professor of bioengineering and mechanical engineering at Stanford University and co-PI of Simbios, are interested in exploring probabilistic approaches. "Probabilistic approaches provide a rigorous method to account for variability between subjects," says Delp.

Viceconti says the deterministic nature of existing models—one input produces one output—really limits their ability to bring the models into clinical practice. "As far as input is good, the output is good. But this is not real life," he says. "If you include a probabilistic approach, you can factor in your ignorance. So you can let a parameter vary widely to see if it matters, and if it doesn't then you can leave it out."

> "Of the dozens of possible parameters that can define the multi-scale phenomenon, which ones really are important and make a difference?" Marco Viceconti says. "That answer will drive the most critical part of the project—not the modeling itself, but the ability to measure in patients the information we need with the accuracy we need."

CONCLUSION

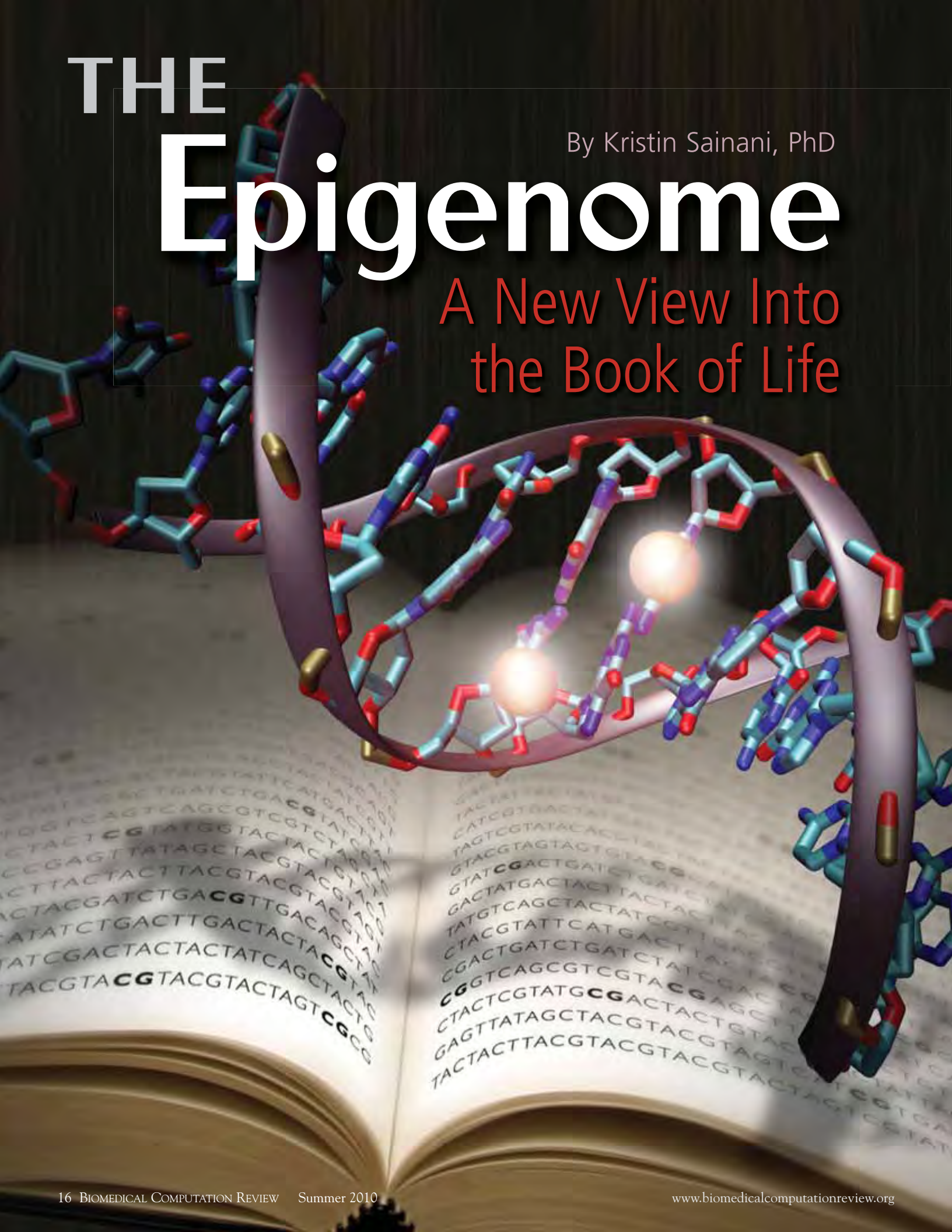
Progress on modeling the physiome reaches well beyond the heart, lung and musculoskeletal examples covered here. Researchers are taking a physiome approach to the kidneys, digestive tract, lymphatic system, and even to some extent the nervous system and the brain, Hunter says.

Before now, Bassingthwaighte says, people have been thinking too narrowly. "But many bright molecular biologists are now trying to be more integrational," he says. "The physiome provides context for that, and for inspiring people to think more broadly." □

THE Epigenome

A New View Into the Book of Life

By Kristin Sainani, PhD



In the early 19th century, Jean-Baptiste Lamarck explained evolution as the inheritance of acquired traits; he believed that changes due to behaviors and exposures in one generation could be passed to subsequent generations. The theory has long since been dismissed. Our actions rarely affect the genetic code of our germline, so our children cannot inherit the consequences, modern genetics assures us.

Surprisingly, however, there may be some truth to Lamarckian inheritance after all. It turns out that our behaviors and exposures can modify our epigenome—causing heritable changes in gene expression without altering the nucleotide sequence. These changes (such as DNA methylation) can be passed down to our offspring, with profound consequences. The phenomenon is well documented in mice, and recent human studies suggest that our food choices and smoking habits may actually affect our kids' and grandkids' risks for diabetes, obesity, and early death.

This is just one of the many potential paradigm shifts arising out of the burgeoning field of epigenetics. Though epigenetics has long been recognized as important—we've known for decades that it is involved in development, cell differentiation, imprinting, and X-chromosome inactivation—it was seen as a side-show to the main attraction, genetics. That view is rapidly changing, however, as there is growing recognition that epigenetics may be just as important as genetics in human health and disease.

In 2008, the NIH launched a $190-million Roadmap Epigenomics Program, which has established centers to map the human epigenome and funded technology development and disease-related projects in epigenetics. In one of the early successes of this initiative, the first complete map of a human epigenome (detailing DNA methylation for two human cell lines) was published in *Nature* last November. Even *TIME* magazine hailed the breakthrough as the number two scientific discovery of 2009. In February of this year, scientists also announced the International Human Epigenome Consortium, a joint effort between the NIH and the European Commission to map 1000 reference epigenomes within a decade.

Studying the epigenome is orders of magnitude more difficult than studying the genome: organisms have a single genome, but hundreds of epigenomes that vary by cell type and developmental stage; the genome comprises just four nucleotides, but the epigenome has many diverse features—including DNA methylation and numerous changes to the proteins that pack DNA into chromatin. The technologies for epigenome-wide studies are just coming online; and they present formidable challenges for computational biologists and bioinformaticians, who must figure out how to process and integrate the enormous amounts of data, as well as correlate them with exposures and diseases.

"The computational epigenetics field is not very developed," says **Christoph Bock, PhD**, a research scholar at the Broad Institute and Harvard Stem Cell Institute. "But this is going to change over the next few years."

Though the field of epigenetics is still in its infancy, the potential payoffs are enormous. Epigenetics has been implicated in cancer, aging, diabetes, mental illness, autism, and Alzheimer's disease. The epigenome is more readily changeable than the genome, which could potentially revolutionize how we prevent, diagnose, and treat disease. Already, several epigenetic drugs are being used to treat cancer.

"The good news is, in terms of future clinical potential, the epigenome is reversible. So, if there's a state that you can alter by chemical means—the methylation profile, for example—you can potentially reverse an epigenetic effect," says **Joseph R. Ecker, PhD**, professor in Genomic Analysis Laboratory at the Salk Institute.

Developing the Epigenetic Toolkit

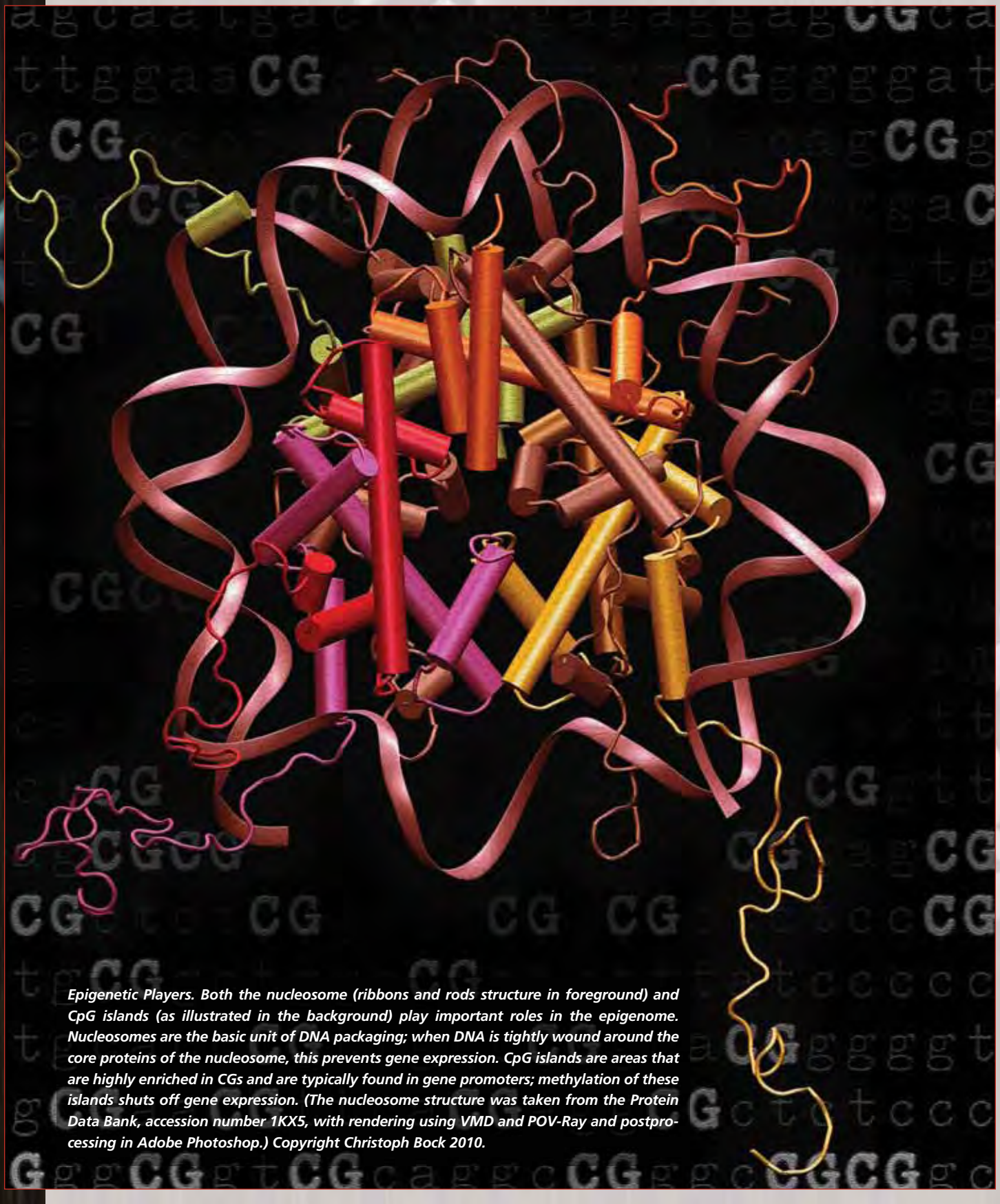
The best studied epigenetic feature is DNA methylation: methyl groups ($-CH_3$) are added to cytosine bases, generally in the context of neighboring cytosines and guanines (CG dinucleotides) such that both DNA strands contain a methyl-C symetrically. Methylation is preserved during mitosis and meiosis, and it serves to silence genes (by blocking transcription factors or recruiting proteins that compact chromatin). Other epigenetic features include biochemical modifications of the histone proteins that wrap DNA into chromatin. For example, adding acetyl or methyl groups to certain lysine residues (e.g. H3K4) in the tails of the histones makes DNA spool more loosely, turning genes on; whereas adding methyl groups to other lysine residues (e.g. H3K9 or K27) makes DNA spool more tightly, shutting genes off. These changes are preserved during cell division, though the mechanisms are not well understood. Non-coding RNAs (RNAs that are transcribed but not made into proteins) also play a role in the epigenome, helping to guide and set up the other epigenetic marks or keeping chromatin open by mere transcriptional activity, says **Michael Zhang, PhD**, professor of computational biology and bioinformatics at Cold Spring Harbor Laboratory (and in the process of moving to University of Texas, Dallas, to set up a new Center for Systems Biology), who is part of the Roadmap initiative to map reference human epigenomes.

The gold standard for detecting methylation is to treat DNA with bisulfite prior to sequencing. Bisulfite converts cytosines to uracils unless they are protected by methylation, so sur-

There is growing recognition that epigenetics may be just as important as genetics in human health and disease.

Opposite: DNA Methylation. This shows a short DNA helix "accgcCGgcgcc" methylated on both strands on the center cytosine. DNA methylation serves to turn genes off by blocking transcription factors or recruiting proteins that compact chromatin. (The structure was taken from the Protein Data Bank, accession number 329D; rendering was performed with VMD and post-processing was done in Photoshop.) Copyright Christoph Bock 2010.

viving Cs represent methyl-Cs. The gold standard for histone marks is chromatin immunoprecipitation, or ChIP: DNA is crosslinked to histone proteins and then exposed to antibodies that recognize specific modifications (e.g., acetylation of lysine 5), followed by microarray analysis (ChIP-on-chip) or direct sequencing (ChIP-Seq). Most epigenome-wide studies to date have been done on arrays, but researchers are increasingly turning to next-generation sequencing (epigenome-wide bisulfite sequencing, ChIP-Seq, and RNA-Seq) in lieu of arrays. Sequencing remains



Epigenetic Players. Both the nucleosome (ribbons and rods structure in foreground) and CpG islands (as illustrated in the background) play important roles in the epigenome. Nucleosomes are the basic unit of DNA packaging; when DNA is tightly wound around the core proteins of the nucleosome, this prevents gene expression. CpG islands are areas that are highly enriched in CGs and are typically found in gene promoters; methylation of these islands shuts off gene expression. (The nucleosome structure was taken from the Protein Data Bank, accession number 1KX5, with rendering using VMD and POV-Ray and postprocessing in Adobe Photoshop.) Copyright Christoph Bock 2010.

cost-prohibitive for large epidemiology studies on human tissue, but even this will change in the next few years.

"The field of epigenetics is moving at an incredible pace, almost exclusively driven by technology development in the sequencing field," says Bock, who tackles computational epigenetics at the Broad Institute, one of the NIH Roadmap's epigenome sequencing centers. "The second generation sequencers are absolutely key for everything we do; and the new machines that are going to become available this year will again change everything we are doing."

"The epigenome-wide methods are advancing so rapidly that if we wait a year, we'll be able to get two to four times as many marks at half the cost," agrees **David A. Bennett, MD**, professor of neurological sciences at Rush University and director of the Rush Alzheimer's Disease Center, who is PI on a Roadmap grant to study the epigenetics of cognitive decline and dementia.

One of the goals of the Roadmap initiative is to help standardize epigenome technology, particularly approaches for processing and analyzing the data.

"At the first Roadmap meeting, it was kind of an eye-opener for me to see how early in the process everyone is—even with cell lines, where they're doing this to known cells types, let alone to a chunk of brain tissue with different cells in it," Bennett says. "It's not just the math and the hardware and the software; there are some big conceptual issues about how to approach some of these datasets."

Bock agrees. "You can just buy the latest Illumina sequencer and download protocols for ChIP-Seq, and experimentally you're fine," Bock says. "But where you're not fine is how you're going to analyze the data."

More and more, he says, that's what his sequencing center needs to provide as a service—the ability to quickly make sense of the data. "Over the last year, the focus was so much driven by just surviving the wave of data that was coming down on us that a lot of the work was algorithmically relatively basic. So there were no complex models involved, but everything had to be ultra-high speed and highly optimized code so we could process these huge amounts of data."

"You can just buy the latest Illumina sequencer and download protocols for ChIP-Seq, and experimentally you're fine," Christoph Bock says. "But where you're not fine is how you're going to analyze the data."

But primary processing of the data is just the first bioinformatics challenge. Researchers must also tackle the higher-level issues: how to integrate the different epigenetic marks with each other and with genome and gene expression data; how to identify and interpret cross-talk between different epigenetic marks; how to make predictions about biological function; and how to compare samples, such as from diseased cases and healthy controls.

"Comparison of epigenomes is not yet a defined problem. We are defining it and implementing solutions as we go," says Aleksandar Milosavljevic.

"Comparison of epigenomes is not yet a defined problem. We are defining it and implementing solutions as we go," says **Aleksandar Milosavljevic, PhD**, associate professor of molecular and human genetics and a PI of the NIH Roadmap's Data Analysis and Coordination Center at Baylor College of Medicine.

Mapping the Human Epigenome

The Roadmap initiative established four Reference Epigenome Mapping Centers, which are covering different aspects of the epigenome (different assays, epigenetics marks, or cell types) with the aim of filling in a matrix of targets. "We are prioritizing the filling of that matrix, so that meaningful analysis can be done at various stages," Milosavljevic says. The first Human Epigenome Atlas data freeze occurred on April 1st.

The mapping centers send their data to the coordinating center (at Baylor), which has developed pipelines for processing and merging the data. "We define and facilitate data flow, data analysis, integrative analysis, and quality control and coordination with all participants," Milosavljevic says. The standards and computational tools that they've developed will also serve as resources for the larger epigenetic community.

One of the first fruits of the Roadmap initiative has been the complete sequencing of a human epigenome at base-level resolution, published in the November 2009 issue of *Nature*. This was a collaborative effort involving some members of the sequencing consortia, led by Joseph Ecker of the Salk Institute. They used bisulfite treatment combined with next-generation sequencing to map the methylation profiles (the "methylome") of a well-known embryonic stem cell line and a differentiated cell line. "Saying you've mapped the human epigenome is correct, but it's not all of the human epigenome—it's just two cell types," Ecker notes. "But it's a start."

Second-generation sequencing was key, Ecker says. "That's what allowed us

to do 30-times coverage of the genome for two different genomes in a fairly short time."

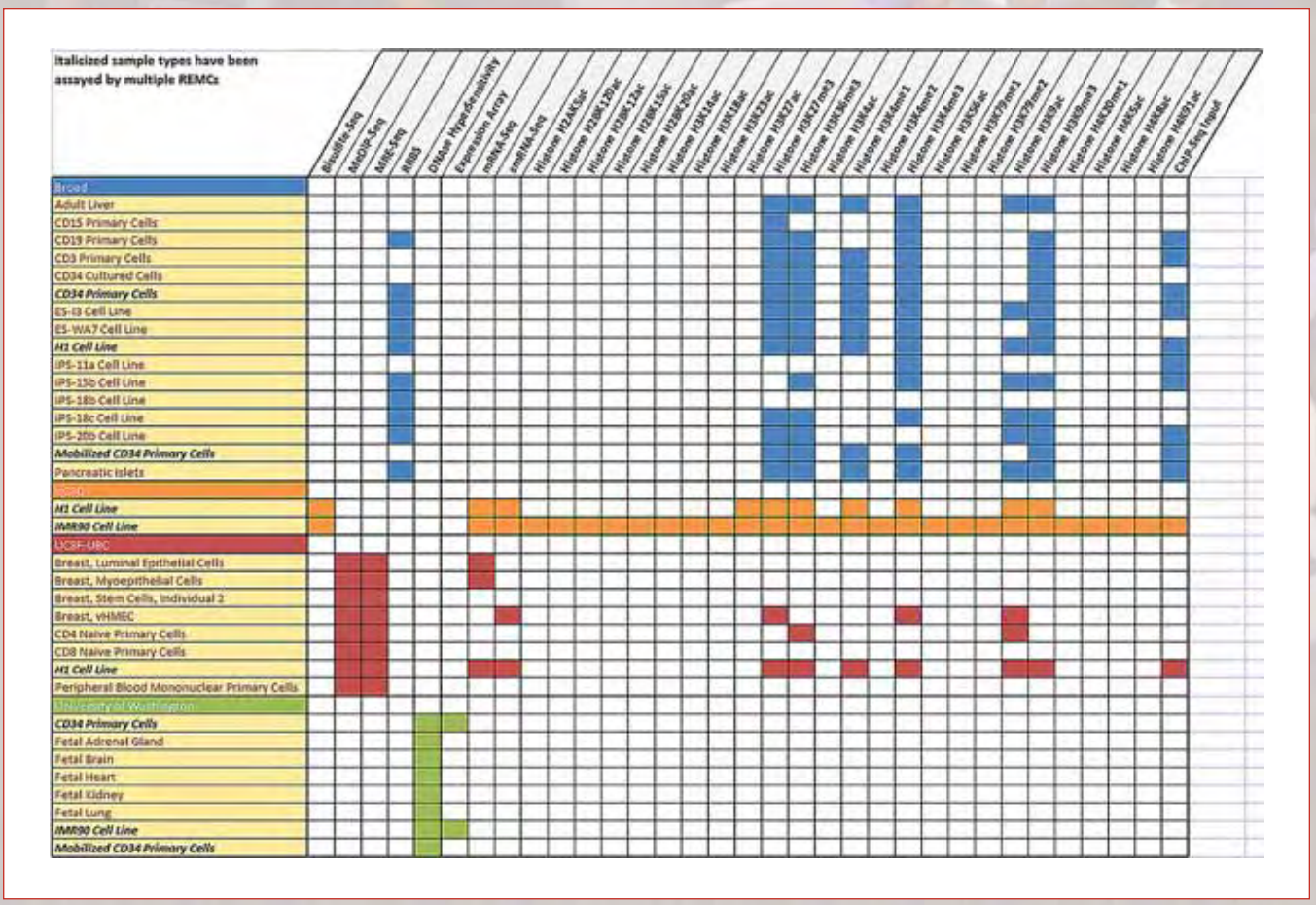
Every step posed major computational challenges, Ecker says. The sequencers produced terabytes-worth of data and just figuring out how to move these data off the machines was initially an issue. They also had to develop new approaches to interpreting the data. For example, they developed an algorithm called "Hammer," which finds methylation sites with a low false discovery rate. "This is the informatics guys' joke because we're measuring methyl-C—also called MC—so it's MC Hammer," Ecker quips.

Previous attempts to look at methylation across the human genome have been array-based. But arrays only look for methylation in certain places. For example, many arrays only query "CpG islands", CG-rich areas of the genome that are typically found in gene promoters. In contrast, whole methylome sequencing reveals methyaltion in all its contexts.

"We didn't have any expectations of what the epigenome was going to look like in terms of its methylation content. So we looked broadly and without bias, and we saw things that were completely unexpected," Ecker says.

Surprisingly, 25 percent of the methylation in the embryonic stem cells was in a non-CG context (i.e, C next to A or C next to T). It is "a complete mystery" how non-CG methylation is maintained during DNA replication, Ecker says. When the DNA unwinds, the complementary portion of the DNA strand contains no Cs (and, therefore, no methyl-Cs) so there's nothing for the methylation machinery to copy on that strand. They also identified another

"We didn't have any expectations of what the epigenome was going to look like in terms of its methylation content. So we looked broadly and without bias, and we saw things that were completely unexpected," Ecker says.



Tackling the Human Epigenome. The NIH Epigenome Roadmap initiative established four mapping centers that are attempting to create a Human Epigenome Atlas by filling in a matrix of targets (different cell lines, assays, and epigenetic features). The first data freeze occurred April 1st, 2010. The samples assayed by the mapping centers are indicated in the rows and the assays in the columns. A live version with viewable and downloadable assay results is available at www.epigenomeatlas.org. Courtesy of Aleksandar Milosavljevic.

novel type of methylation: differentiated cells, but not stem cells, contain long stretches of half-methylated DNA (which they called "partially methylated domains"). The significance of these domains is unknown, but other groups have now identified them in cancer cells, Ecker says.

"I think we learned from this that we should probably have less bias about what we're going to find, otherwise we won't find it," Ecker says.

Connecting Nurture with Nature

Part of the excitement surrounding the epigenome is that it is far more responsive to the environment than the genome. In fact, epigenetics blurs the lines of the nature versus nurture debate, as it turns out that our environments can extensively impact our inherent biology. Monozygotic twins start life with highly similar epigenomes, but their epigenomes diverge as they age, particularly if they've had different environmental exposures.

"Epigenetics gives you a way to bridge the gene versus environment question. It is genetic in a way, because you can measure it just like DNA, but it is also influenced by the environment," Bock says. "So you can't really give people a machine to carry with them that over a lifetime measures all environmental exposures. But the epigenome might provide such a machine, because it responds to all kinds of external influences." Epigenetic changes have been related to aging, smoking, diet, alcohol, asbestos, arsenic, inflammation, heavy metals, ultraviolet radiation, infection, toxins, stress, and psychological abuse.

"I think the sexiness of epigenetics is that there's this potential connection between exposures and the biological consequences of them. You can even measure the biological consequences of social impacts," says **Margaret Daniele Fallin, PhD**, associate professor of epidemiology at the John Hopkins University Bloomberg School of Public Health, who is co-PI on a Roadmap grant to study the epigenetics of autism.

The reference epigenomes will be incredibly valuable, but they won't tell us much about how epigenomes vary in healthy populations and in response to the environment, says **Karl Kelsey, MD**, professor of community health at Brown University. Conducting such epidemiological studies is a much harder task because you have to study many people and many different tissues, some of which may not be easily accessible (unlike for genetic studies, where DNA can be acquired from any accessible cells).

"Every tissue in your body has a different epigenome. So if you're looking for variability, you've got to look at every tissue, and suddenly the problem becomes more complicated," Kelsey says. Most studies to date have used array-based technologies, because it's still cost-prohibitive to sequence so many samples.

For example, in a 2009 paper in *PLoS Genetics*, Kelsey and his colleagues used Illumina GoldenGate arrays—which probe methylation in 1500 targeted CpG sites from hundreds of genes—to characterize 217 normal human tissue samples from 10 anatomical sites. One of the computational challenges is that individual methylation events cannot be assumed to be independent, so statistical methods need to account for this correlation across methylation sites. "I think this is a poorly understood and poorly recognized problem, for methylation data certainly," Kelsey says. They used a particular recursive partitioning algorithm (developed by **E. Andrés Houseman, ScD**, assistant professor of community health at Brown University) that uses mixture models to deal with this problem. "It's a very interesting solution," Kelsey says.

They showed that methylation increases with age within CpG islands, but decreases with age at other loci. In previous work, they showed that smoking methylates the promoter region of the p16 tumor suppressor gene in lung tissue. These same methylation changes have also been linked to cancer.

Diagnosing and Treating Cancer

Epigenetics has been a hot topic among cancer researchers for more than a decade—much longer than for other diseases. Epigenetic changes have been identified in almost all cancers. Indeed, according to Kelsey, "Epigenetics is an equal partner to genetic change in creating cancer." The epigenome can silence tumor suppressor genes or wake up oncogenes or imprinted genes.

Epigenetics holds promise for early detection, as many of the changes appear to occur early in the progression to cancer. "In many cases, before tumor suppressor genes get deleted, their promoters may first get shut off by DNA methylation," Zhang says. Cancer cells often slip into the blood, so it may be possible to pick up epigenetic signatures from a simple blood test. "So, people have done a lot with looking at changes in blood, urine, and sputum," Kelsey says. "And I think that's a real possibility. To date, most of the assays haven't really borne a lot of fruit. But we're just really starting to apply them."

Beyond diagnosis, epigenetic markers may be able to predict the development of cancer even before it appears. For example, about 1 in 200 patients with Barrett's esophagus, a premalignant condition, will go on to develop esophageal cancer each year. Currently,

"I think the sexiness of epigenetics is that there's this potential connection between exposures and the biological consequences of them. You can even measure the biological consequences of social impacts," says Margaret Daniele Fallin.

there is no way to predict who will get cancer, so every patient must undergo repeated endoscopies, which entail high costs, inconvenience, and anxiety, says **Stephen J. Meltzer, MD**, the Hendrix/Myerberg Professor of Medicine and Oncology at John Hopkins University.

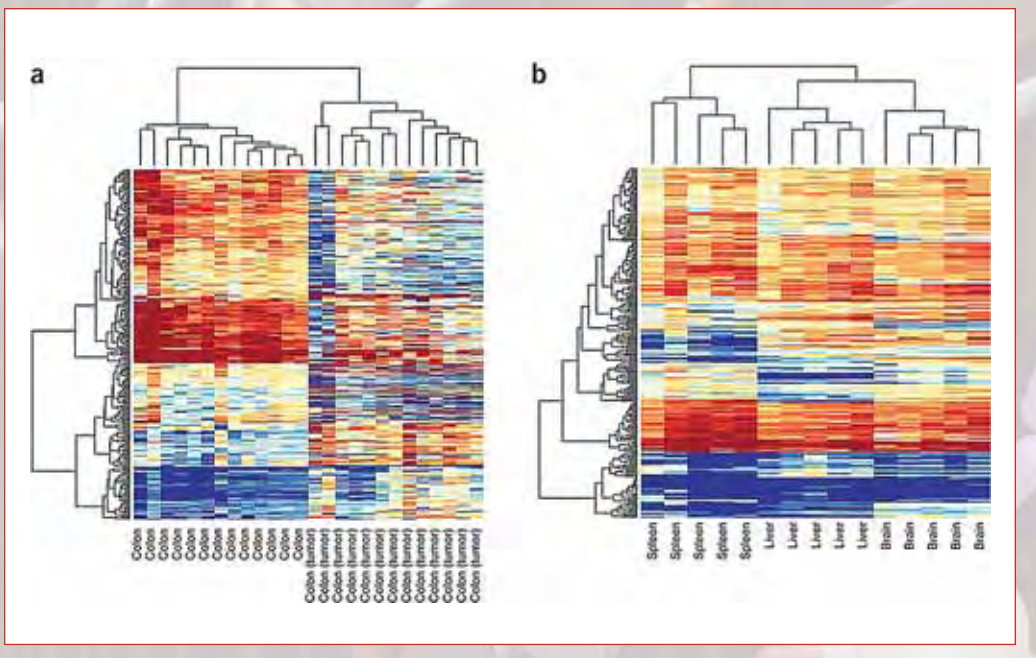
His team has identified a panel of eight epigenetic markers (hypermethylated tumor suppressor genes) that can correctly distinguish progressors from nonprogressors about 75 percent of the time. "When we developed this assay, we didn't have the genome-wide or the epigenome-wide tools that are now available. So some of the markers we originally chose to study in-depth may not be the best or the only ones we end up with," says Meltzer, who has a Roadmap grant to search the epigenome for novel markers.

Epigenetic changes may be reversible, which makes them a prime target for treatment and prevention. Already, four epigenetic drugs that re-activate silenced genes (presumably tumor suppressor genes) have been approved for treating blood cancers: two that demethylate DNA and two that maintain histone acetylation (histone deacetylase inhibitors). These drugs lack specificity, and in theory could inadvertently turn on oncogenes. But, so far, they appear to do more good than harm. Researchers are hopeful that new agents and combinations of agents will be even more effective and work on solid tumors.

Epigenetic cancer research to date has focused on the most obvious targets—methylation of tumor suppressor genes or CpG islands. But the most relevant epigenetic events may be happening outside of these contexts, as demonstrated by research at Johns Hopkins University. "Andy [Feinberg] had an intuition that that wasn't the place to be looking. I'm not quite sure why, but it turned out to be right. He wanted an array that didn't bias toward genes or islands. We spent a lot of time developing that array," says **Rafael Irizarry, PhD**, professor of biostatistics at the Bloomberg School of Public Health, who collaborated on the array—called CHARM (comprehensive high-throughput arrays for relative methylation)—with **Andrew Feinberg, MD**, professor of molecular medicine, oncology, and molecular biology & genetics at Johns Hopkins University School of Medicine. The array probes regions of high CG content regardless of whether they are CpG islands.

Figuring out how to analyze the data from the arrays has been a challenge, Irizarry says. Just as with gene expression arrays, there is a multiplicity problem—many signals will arise simply by chance rather than as the result of true differences between cases and controls; and distinguishing these is tricky. "It's the same exact problem, except it's harder," Irizarry says. Gene expression arrays focus on predefined units, genes, whereas methylation arrays focus on open-ended regions; so there is uncertainty in defining the regions of interest as well as interpreting the intensity of their signals. "Basically, what it comes down to is there are really two dimensions. It's not just the level of expression; it's the size of the region and the height," Irizarry says. They are still fine-tuning their algorithms, he says.

In a 2009 paper in *Nature Genetics*, Irizarry and colleagues used CHARM to compare normal tissue and colon cancer samples. As expected, they found many differences in the methylation profiles. What was surprising is that these differentially methylated regions were rarely in CpG islands; rather, they were in regions adjacent to the islands, which they deemed "shores." Interestingly, while CpG islands generally become methylated during cancer, the shores (which may represent alternative transcription start sites) were equally likely to become demethylated as methylated. Their findings imply that many cancer studies to date have been looking in the wrong places for key methylation changes. Irizarry's team is now making the switch from arrays to next-generation sequencing. "That introduces a whole new set of problems," Irizarry says.



Cancer's Epigenetic Signature. The methylation pattern of colon cancer is distinct from normal colon tissue and other tissues in the body. Certain regions are hypermethylated (red) and others are hypomethylated (blue). The tissues were analyzed using the CHARM assay, followed by unsupervised hierarchical clustering—which perfectly differentiated cancer from non-cancer and colon tissue from other tissue types. Reprinted by permission from Macmillan Publishers Ltd: Irizarry et al., The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. **Nature Genetics** ***41:178-86 (2009).***

Epigenetic changes beyond methylation also play a key role in cancer, but these haven't been well studied, says **Terumi Kohwi-Shigematsu, PhD**, senior scientist in the Life Sciences Division of the Department of Energy's Lawrence Berkeley National Laboratory, who is PI on a Roadmap grant to study the epigenomics of breast cancer. Her work focuses on higher-level changes in chromatin structure that affect gene

expression. "Epigenetics is probably governed at these higher architectural levels," she says.

In a 2008 paper in *Nature*, Kohwi-Shigematsu's team reported that the protein SATB1, which helps fold chromatin, is overexpressed in aggressive breast cancer cells and correlates with metastasis and poor survival. SATB1 binds to certain regions of DNA and recruits histone-modifying and chromatin-remodeling enzymes that in turn alter the expression of about 1000 genes, many of which are known players in metastasis. "SATB1 provides the mechanism for assembling all this 3D genome architecture to locally determine epigenomic modifications and thereby regulate a large number of genes," Kohwi-Shigematsu says. Introducing the protein into non-metastatic breast cancers in mice induces invasive tumors while depleting it from metastatic cells in mice reverses tumors. Thus the protein has implications for both prognosis and treatment. SATB1 binds to certain specialized regions of DNA throughout the genome. Kohwi-Shigematsu's team will use ChIP-Seq to map these areas and identify which ones are bound by SATB1 in aggressive breast cancer. Computational epigenetics researchers are still working out the optimal algorithms for analyzing ChIP-Seq data.

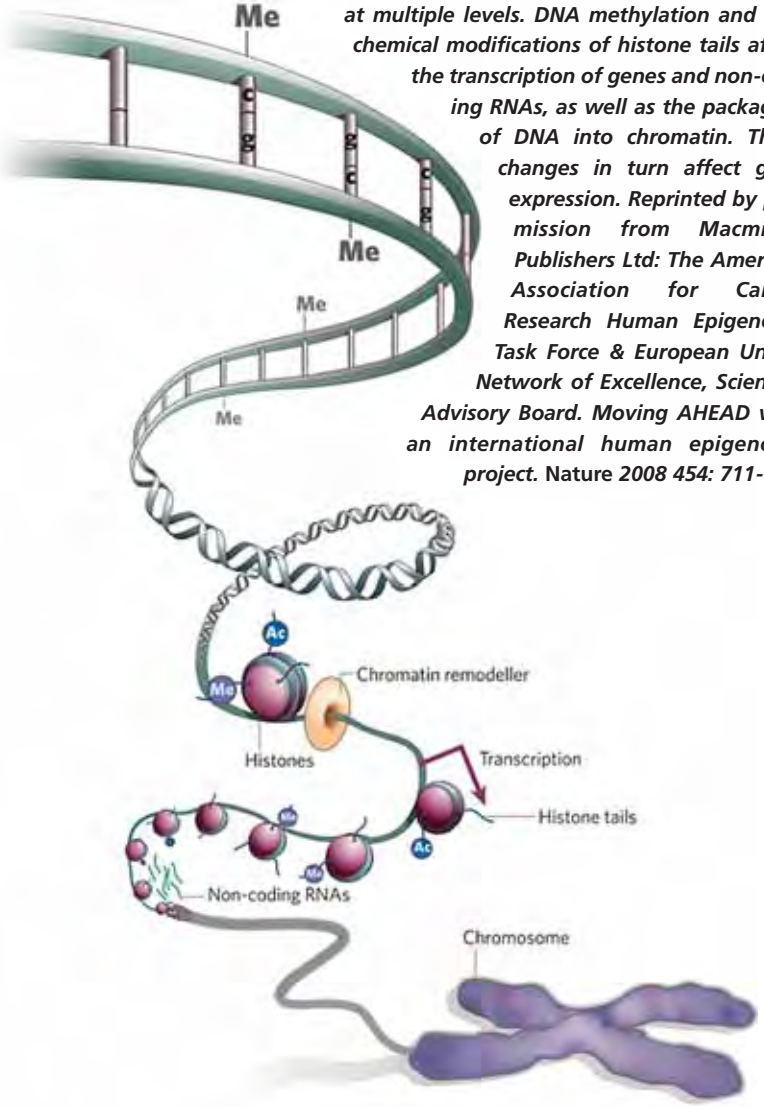
Understanding the Brain

Methylation plays a key role in brain development; for example, several developmental disorders involve loss of genomic imprinting and one (Rett Syndrome) is caused by a mutation in the enzyme that methylates DNA. Epigenetics has also been implicated in mental illness, and several psychiatric drugs have known epigenetic effects. For example, valproic acid, used to treat epilepsy and bipolar disorder, is a histone deacetylase inhibitor.

"I think epigenetics is a lot more exciting than the genome, especially for the brain, which is so plastic," Bennett says.

Several pivotal epigenetic studies have focused on the brain. For example, in a groundbreaking paper in *Nature Neuroscience* in 2004, scientists from McGill University showed that maternal nurturing directly affects psychological development through an epigenetic mechanism. When infant rats were neglected by their mothers (whether biological or foster mothers), the glucocorticoid receptor gene in their brains became methylated (in the promoter). This change persisted into adulthood and caused the rats to be highly stressed. In contrast, baby rats that were extensively groomed and cared for by their mothers had reduced methylation of the gene and became more relaxed adults. Moreover, both effects were reversible in the adults—nurtured rats given a shot of methionine (a DNA-methylating agent) to the brain became stressed and neglected rats given a histone deacetylase inhibitor (which indirectly reduces methylation) became calm.

There's mounting evidence that epigenetics is a key player in autism as well, Fallin says. She received a Roadmap



Layers of the Epigenome. The epigenome acts at multiple levels. DNA methylation and biochemical modifications of histone tails affect the transcription of genes and non-coding RNAs, as well as the packaging of DNA into chromatin. These changes in turn affect gene expression. Reprinted by permission from Macmillan Publishers Ltd: The American Association for Cancer Research Human Epigenome Task Force & European Union, Network of Excellence, Scientific Advisory Board. Moving AHEAD with an international human epigenome project. **Nature** ***2008 454: 711-15.***

"I think epigenetics is a lot more exciting than the genome, especially for the brain, which is so plastic," David Bennett says.

grant to study epigenetics in the EARLI (Early Autism Risk Longitudinal Investigation) cohort. The study recruits participants when they learn they're pregnant and follows them through the birth of the baby and the first three years of life. "So we get the whole window of potential exposures, and then we get the early development of the child," Fallin says. Her team will correlate exposures with methylation changes in the mothers' and babies' blood cells and then try to link these to autism outcomes.

Currently, they are using arrays to study the methylation profiles, but "just like everyone else, we are thinking about how you get this directly from sequencing," says Fallin, who works closely with Irizarry.

At the other end of life, epigenetics may play a role in Alzheimer's disease and dementia. "When I came across the epigenetics literature, investigators were just beginning to conduct preclinical animal studies examining how the brain might be using epigenetic marks as a way of coding long-term memory," Bennett says. He received a Roadmap grant to look at epigenetics within two long-standing studies of older people: the Rush Memory and Aging Project and the Religious Order study. Participants undergo annual cognitive testing, provide information about life experiences, and, when they die, donate their brains to the study (nearly 800 brains so far). Bennett's team will obtain methylation profiles for brain tissue using next-generation sequencing. They have also received an ARRA stimulus grant to collect data on histone modifications.

"The idea is to link epigenetic changes initially to cognitive phenotype and then to psychological and experiential factors," Bennett says. "Subsequently, we'll be able to bring in the genome-wide data, because the effects may vary by genetics, kind of in the background."

"Putting all these data together is not straightforward," Bennett says. "In the grant, we're doing our best to describe our approach. But I think it's really unclear until the data are there. Certainly, it's so new for data from human brains."

Vindicating Lamarck

Studies in mice show that epigenetic changes can be passed down through multiple generations. Agouti mice carry a mutated gene that gives them a yellow coat and a propensity for diabetes and obesity. But when pregnant Agouti mice are fed extra methionine and folic acid—nutrients involved in DNA methylation—their children turn out brown, lean, and healthy; they still carry the defective gene, but it has been silenced through methylation. When these offspring reproduce, they pass the silenced gene to their children regardless of their diets; thus, the grandmother's diet determines the grandchildren's phenotypes.

Heritable epigenetic changes can occur at other times in the life cycle as well, not just during fetal development. In a 2009 paper in the *Journal of Neuroscience*, when adolescent mice with a genetic defect in memory were exposed to an enriched environment (with novel objects, social interaction, and voluntary exercise), their memories improved. "Our original observation was that an enriched environment can overcome the biochemistry of a genetic defect by opening up a new signaling pathway," says **Larry Feig, PhD**, professor of biochemistry and neuroscience at Tufts University School of Medicine, who led the research. Unexpectedly, when the enriched mice reproduced (females only), their children also had improved memories even though they were never exposed to the stimulating environment. This was true even when the children were raised by foster moms with poor memories.

"It was a surprise. It wasn't an area of research that we were actively working on," he says. "But when I went back to the literature and looked more carefully, there were growing examples of epigenetic transgenerational inheritance. So it wasn't as farfetched as my initial thoughts."

Feig's team has not yet identified the epigenetic change responsible. "So this is an example of epigenetics by all assumptions, but we don't have any details yet to pin it down," he says. They are collaborating with researchers at the Broad Institute on epigenome-wide studies to search for the relevant changes. Feig also received an NIH Challenge Grant to study whether adolescent mice exposed to negative experiences, such as smoking and stress, also undergo heritable epigenetic changes.

Epigenetic inheritance is more difficult to study in people, but recent studies suggest that it does occur. In a series of studies from Europe, researchers collected multi-generational data from an isolated community in Northern Sweden that experienced alternating periods of plenty and famine throughout the 19th century. They showed that the paternal grandsons of men who were exposed to periods of overabundant food before puberty (a key stage for sperm development) were at increased risk of diabetes and early death. The paternal granddaughters of women who were exposed to abundant food in utero or during infancy (key stages for egg development) had increased mortality. The pattern of inheritance suggests that the epigenetic effect may be occurring on sex-linked genes, though the epigenetic mechanism has yet to be definitively proven.

If Lamarckian inheritance turns out to be a real phenomenon in people, this will be both an empowering and daunting shift in how we think about evolution and the destiny of our descendants.

If Lamarckian inheritance turns out to be a real phenomenon in people, this will be both an empowering and daunting shift in how we think about evolution and the destiny of our descendants. As scientists continue to probe the largely unexplored territory of the epigenome, this promises to be just one of many surprises. □

Under TheHood

BY PAT HANRAHAN, PhD, PROFESSOR OF COMPUTER SCIENCE AT STANFORD UNIVERSITY

Using Domain Specific Languages to Access Parallel Computing in All Its Forms

LINDA A. CICERO / STANFORD NEWS SERVICE

Increasingly, parallel computing is becoming pervasive. The processors found in laptop and desktop machines now have two to four cores. Cheap and widely available graphical processing units (GPUs) contain hundreds, and soon thousands, of cores. And with the advent of cloud and grid computing infrastructure, every scientist can access thousands of processors and petaflops of computing.

For scientists who want to do simulation or data analysis, this is great news. Unfortunately, writing parallel programs is hard. The processor on your desktop uses a different programming model (e.g. threads and locks) than a GPU (that uses CUDA or OpenCL) or a cluster (that uses MPI). Because hardware designers want their computers to be efficient, they specialize them for different types of workloads. A GPU, for example, is ten times more power efficient than a CPU at floating point intensive, high-throughput calculations. This need for power efficiency will lead to even more heterogenous machines in the future. Learning the low-level details of how to program many different types of machines efficiently takes a long time. And computational biologists have other fish to fry—solving biological research problems.

I believe that to take advantage of emerging parallel computers without learning how to program different machines, computational biologists will need to write software at a higher-level: They will need to use domain-specific languages and libraries (DSLs). A DSL is an environment that is tailored to a particular domain or task. Most scientists already use DSLs such as matlab, R, and latex routinely in their work. DSLs take the grunge out of programming, letting the computational scientist focus on the science, not on the computer hardware.

Unfortunately, DSLs have a reputation for being slow (often because they are interpreted). This need not be the case. Two widely used DSLs are OpenGL and SQL.

OpenGL is a graphics library that can be thought of as a DSL for graphics. OpenGL drivers implement graphics commands on highly parallel GPUs in an efficient way.

SQL is the *lingua franca* of databases. Because queries are expressed at a high-level, it is possible to implement SQL on large datacenters. These two DSLs are both portable and efficient.

Because DSLs operate at a higher-level of abstraction, they can often be automatically parallelized. Take for example molecular dynamics or n-body algorithms. There are very efficient n-body algorithms that are tailored to different types of parallel machines. A DSL for molecular dynamics can build these algorithms into the system, and thus a program written in the molecular dynamics DSL will run portably on different types of parallel computers. By contrast, a general-purpose parallelizing compiler could never automatically discover these algorithms, if the molecular dynamics application is written in a lower-level programming language like C++.

I believe that to take advantage of emerging parallel computers without learning how to program different machines, computational biologists will need to write software at a higher-level: They will need to use domain-specific languages and libraries (DSLs).

As computational biology evolves and new software is developed, we need to do two things to ensure that this software will run on the parallel computers of the future. First, we need to assemble teams of computational biologists and computer scientists to develop DSLs for the major areas of computational biology. One example of successful collaborations of this type is the OpenMM Project developed by Simbios at Stanford. DSLs could be developed for many other areas including finite element calculations, fluid flow, machine learning, and data analysis, to name a few. Second, computer scientists need to build tools that make it easier to build DSLs. Currently, programs like Matlab are implemented from the ground up. DSL writers essentially "roll-their-own" systems. We need a general infrastructure so that DSLs can be easily built and extended by small groups of people. If we succeed, future computational biologists will have access to extraordinary computing capabilities, which in turn will enable many scientific discoveries. □

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

Biomedical Computation Review
Simbios An NIH National Center for Biomedical Computing
Stanford University
318 Campus Drive
Clark Center Room S231
Stanford, CA 94305-5444

seeing science

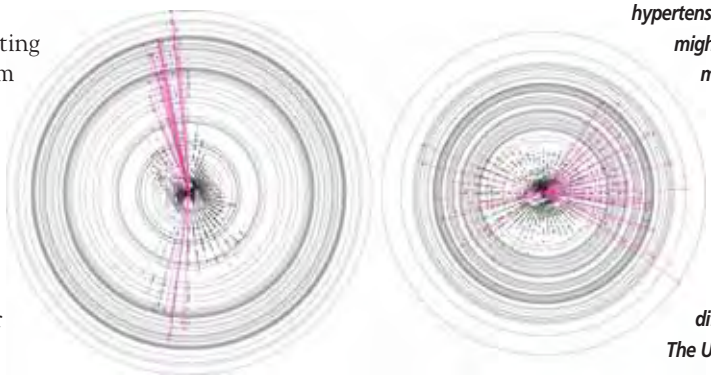
SeeingScience

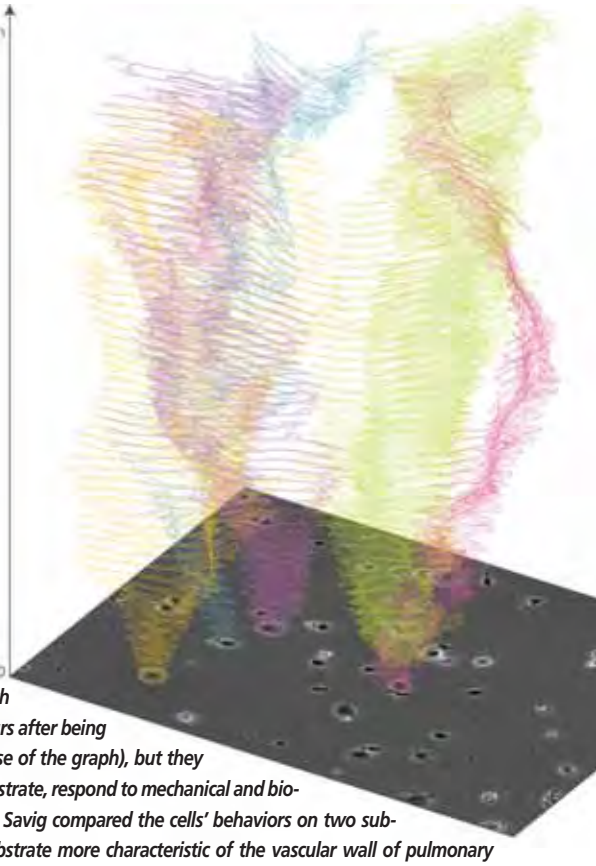
BY KATHARINE MILLER

Architectural Computation Visualizes Cell Choreography

Several years ago, **Jackie Wong, MArch**, mentored by **Jenny E. Sabin, MArch**, an architectural designer and lecturer in the School of Design at the University of Pennsylvania, developed a tool for understanding and visualizing ice dancers' movements. He mapped the relationships between the arms, legs and head of the skater to generate visual patterns that describe the structure of various choreographies. Using this architectural work as inspiration, **Erica Savig, MArch** (UPenn 2008), now a graduate student in cancer biology at Stanford, and **Mathieu C. Tamby, PhD**, a post-doctoral fellow at UPenn, devised a related algorithm to analyze and understand how the tissue microenvironment within pulmonary arteries alters the movement of vascular smooth muscle cells in the context of pulmonary hypertension. The ongoing work is part of a collaboration between architects and cell biologists at UPenn known as **Sabin+Jones LabStudio**, which was founded and is co-directed by Sabin and **Peter Lloyd Jones, PhD**, associate professor of pathology and laboratory medicine at Penn, (and now lecturer in Architecture).

"Jackie Wong had existing dance steps and visualized them into 3-D representations," says Savig, one of the first architectural students recruited into the unique collaboration, "We worked backwards, visualizing cell movements to search for unseen patterns and the fine details of their unknown choreographies." □





Tracing Cell Choreography to Determine How Microenvironment Alters Cell Behavior. *This colorful 3-D graph traces the morphologies and movements of five different smooth muscle cells through time (vertical axis). Two hours after being seeded, the cells are small and nearly round (base of the graph), but they soon spread their filopodia as they probe the substrate, respond to mechanical and biochemical signals, and interact with one another. Savig compared the cells' behaviors on two substrates: fibrillar and non-fibrillar collagen—a substrate more characteristic of the vascular wall of pulmonary hypertensive patients. To identify visual signatures that might one day help personalize diagnosis of pulmonary hypertension, she created visual abstractions of the tracings. Here looking at the data from two cells compressed into 2-D (at left), the disorder of cell movements on non-fibrillar collagen (rightmost image) is clearly visible. The algorithms behind these computationally generated representations allow extraction of numerical measurements for these differences. Courtesy of Sabin+Jones LabStudio, The University of Pennsylvania.*