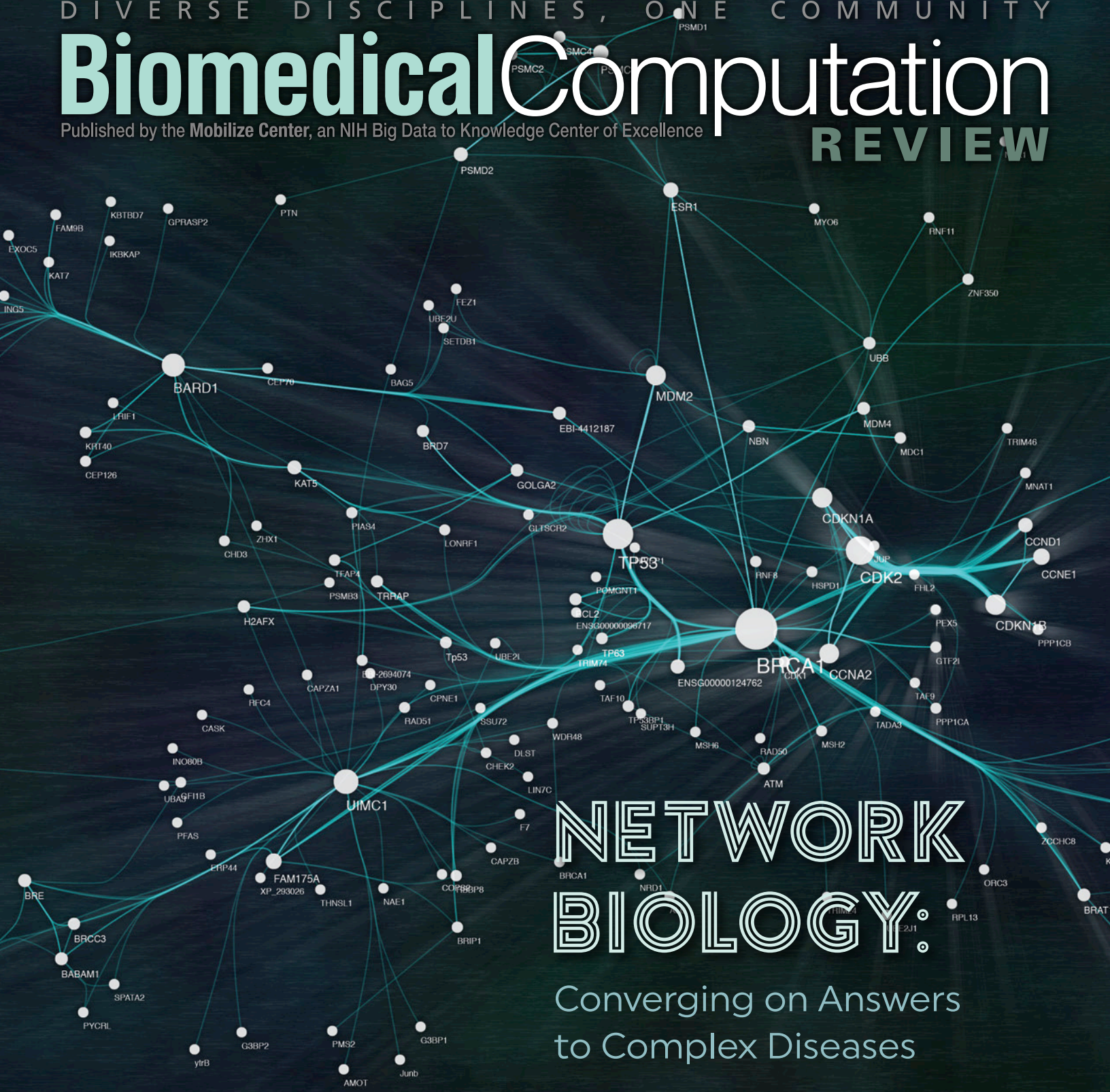
DIVERSE DISCIPLINES, ONE COMMUNITY

Biomedical Computation REVIEW

Published by the **Mobilize Center**, an NIH Big Data to Knowledge Center of Excellence



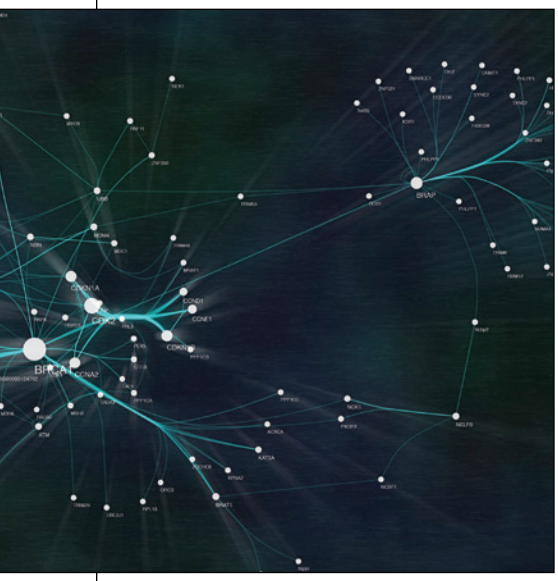
NETWORK BIOLOGY:

Converging on Answers to Complex Diseases

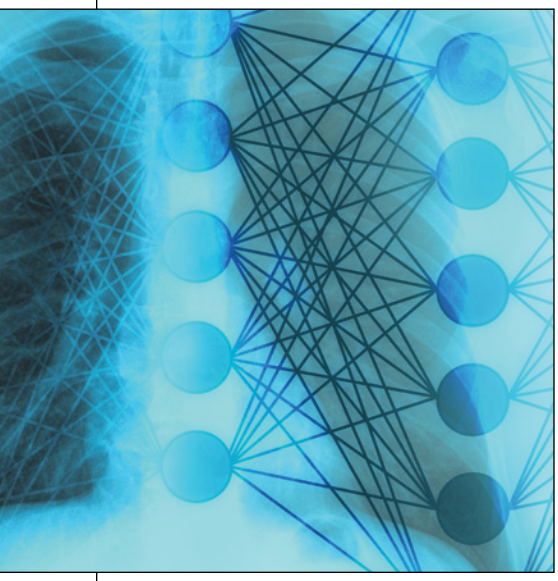
PLUS:
Deep Learning and the Future of Biomedical Image Analysis

WINTER 2017

Contents Winter 2017



14
Network Biology:
Converging on Answers to Complex Diseases
BY KRISTIN SAINANI, PhD



22
Deep Learning
And the Future of Biomedical Image Analysis
BY ALEXANDER GELFAND

Cover Art and Pages 14-15: Created by Rachel Jones of Wink Design Studio using: Network image supplied under Creative Commons License, created by Keiichiro Ono.

Pages 22-23: Created by Rachel Jones of Wink Design Studio using: Lung image © Kravka | Dreamstime.

DEPARTMENTS

1 GUEST EDITORAL
NETWORK BIOLOGY—AT THE FRONTIER
BY MARK CRAVEN, PhD

2 MOBILIZE NEWS
THE "LEARNING TO RUN" CHALLENGE:
BY KATHARINE MILLER

UNDERCURRENTS

3 IS CLINICAL GENOMICS TESTING WORTH IT?
BY KATHARINE MILLER

6 PROTEINS FOR EVERY OCCASION
BY MARCUS WOO

9 BIG DATA FEEDS UNDERSTANDING OF OBESITY
BY ALEXANDER GELFAND

12 CAN BIG DATA STOP OPIOID ABUSE?
BY SARAH WILLIAMS

29 UNDER THE HOOD
WHAT VALUE COULD FRACTALS ADD TO BIOMEDICAL IMAGE ANALYSIS?
BY CHAND T. JOHN, PhD,

30 SEEING SCIENCE
AN AUTOMATED SUPERTREE:
A Model for Extracting Literature-based Knowledge
BY KATHARINE MILLER

Winter 2017
Volume 13, Issue 2
ISSN 1557-3192

Co-Executive Editors
Scott Delp, PhD
Russ Altman, MD, PhD

Associate Editor Joy Ku, PhD

Managing Editor Katharine Miller

Science Writers
Alexander Gelfand, Katharine Miller, Kristin Sainani, PhD, Sarah Williams, Marcus Woo

Community Contributors
Mark Craven, PhD, Chand T. John, PhD

Layout and Design
Wink Design Studio

Printing
AMP Printing

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

Office
Biomedical Computation Review
Stanford University
318 Campus Drive
Clark Center Room W352
Stanford, CA 94305-5444

Publication is supported by NIH Big Data to Knowledge (BD2K) Research Grant U54EB020405.

Information on the BD2K program can be found at http://datascience.nih.gov/bd2k.

The NIH program and science officers for the Mobilize Center are:

Grace Peng, National Institute of Biomedical Imaging and Bioengineering,

Theresa Cruz, National Institute of Child Health and Human Development,

Daofen Chen, National Institute of Neurological Disorders and Stroke

***Biomedical Computation Review* is published by:**

The Mobilize Center,
an NIH Big Data to Knowledge (BD2K) Center of Excellence

mobilize.stanford.edu

mobilize
Center for Mobility Data Integration to Insight

BY MARK CRAVEN, PhD
Professor of biostatistics and medical informatics at the University of Wisconsin, Madison, and principal investigator of the Center for Predictive Computational Phenotyping, a Big Data to Knowledge Center of Excellence

NETWORK BIOLOGY—AT THE FRONTIER

Living systems are characterized by intricate networks of interactions among tens of thousands of entities within cells and across extracellular milieus. The entities in these networks include such things as genes, proteins, and small molecules. The interactions describe relationships such as transcriptional regulation, posttranslational modification, and complex formation, among others.

Our understanding of phenomena like disease states, responses to toxins, and basic biological processes such as circadian rhythms and cellular differentiation, hinges on our ability to characterize and make inferences about these networks. This challenge is daunting given that the processes defined by the networks are dynamic; the relevant entities and interactions in the networks vary across cell types and contexts; and our knowledge of the entities and interactions is incomplete even for the simplest, single-cell organisms. Nevertheless, substantial progress is being made in developing computational methods that augment our understanding of the biological networks underlying processes, responses and states of interest. As highlighted in this issue, computational network biology is rapidly advancing with innovations on several important fronts.

Addressing incompleteness. The gaps in our knowledge of intracellular networks are being partially filled in by novel technologies for more thoroughly identifying specific types of interactions such as the protein-protein interactions that are being detected in the BioPlex project at Harvard. An alternative strategy to handling incompleteness is to screen for genetic interactions using knockout or knockdown methods. Novel algorithms are being developed to make inferences about how sets of gene products interact based on the results of these genetic-interaction screens.

Incorporating multiple types of interactions. Although useful inferences can sometimes be made by analyzing networks composed from a single type of interaction (e.g. protein-protein interaction or gene co-expression), we can clearly gain higher fidelity representations of biological processes by constructing and reasoning with network models that incorporate multiple types of interactions.

Network descriptions of diseases and patients. Whereas early network models focused on routine processes in model organisms, recent research has demonstrated that network models can provide insight into diseases as varied as autism, breast cancer, Parkinson's (described in this issue), and viral infections.[1] Moreover, many of the algorithms that have been applied to intracellular networks can be applied to other types of networks, such as patient similarity networks in which the nodes represent patients and the edges represent phenotypic similarity.

Exploiting relationships among organisms and cell types. Gaps in our understanding of networks in one type of cell can often be alleviated by taking advantage of information from related cells.[2] For example, the TransposeNet method described in this issue has aided in the characterization of relationships among Parkinson's risk genes by mapping a relevant subnetwork from yeast onto an inferred human network.

In addition to these avenues of research, there are other directions where we can expect to see significant innovations in the near future. One important challenge is to devise network models that provide more expressive and faithful representations of the underlying biology. Such models will incorporate representations of epigenetics, cellular compartments, spatial and transport relationships, intercellular interactions, and host-microbiome interactions, among other aspects.
A second area that is ripe for further exploration entails approaches that specify how network responses change as a function of genetic variation and environmental exposures. Another promising area: algorithms for optimally selecting the most informative experiments to refine network models.[3] And a fourth key direction is devising network models that span multiple scales, from molecules to whole organisms and their microbiomes. Such network models hold the promise of capturing in substantial detail how patient-level descriptors like symptoms and diseases are manifested all the way down to the molecular level, thus helping to drive advances in precision medicine. □

References

1. Integrated systems biology analysis of KSHV latent infection reveals viral induction and reliance on peroxisome mediated lipid metabolism. Sychev ZE, Hu A, DiMaio TA, Gitter A, Camp ND, Noble WS, Wolf-Yadlin A, Lagunoff M. *PLoS Pathogens* 13(3):e1006256, 2017.

2. Inference of cell type specific regulatory networks on mammalian lineages. Chasman D, Roy S. *Current Opinion in Systems Biology* 2:130-139, 2017.

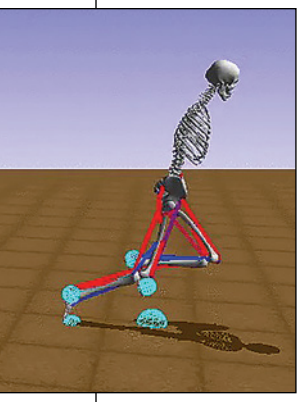
3. A review of active learning approaches to experimental design for uncovering biological networks. Sverchkov Y, Craven M. *PLoS Computational Biology* 13(6):e1005466, 2017.

BY KATHARINE MILLER

THE "LEARNING TO RUN" CHALLENGE:

Engaging Data Scientists in Biomechanics

Most of us take walking or running for granted. But injury or neurological disease can cause these basic skills to deteriorate to such a degree that they need to be re-learned. To better understand how the brain accomplishes such learning, Stanford University's Mobilize Center challenged the research community to a competition: develop a controller (essentially a brain) that will allow a physiologically-based human model to navigate through a complex obstacle course as quickly as possible. At the December 2017 Neural Information Processing Systems (NIPS) conference, winners of the "NIPS 2017: Learning to Run" challenge will be announced.



The nature of the challenge has drawn many participants who are more familiar with data science concepts—such as reinforcement learning—than they are with muscles and bones, says **Łukasz Kidziński, PhD**, the Mobilize Center postdoctoral student who co-organized the competition in collaboration with researchers at the University of California, Berkeley, and École polytechnique fédérale de Lausanne in Switzerland. "Because they come at the problem from a different field, they use other types of skills to find new ways to solve challenging biomechanical problems," he says.

For the competition, Kidziński says, "We gave people a model with muscles and bones and interactions and constraints of the human body, and they had to build a brain for this body." This simplified lower body model, which consists of bones and 18 key muscles, was designed in OpenSim, the biomechanical simulation platform developed by Mobilize Center researchers. "Other experiments had shown that these are enough muscles to synthesize human-like gait," Kidziński says. He and his colleagues also set some reasonable constraints on the model, such as limits on the forces that muscles can exert, and a requirement that the model can only go forward. In addition, they built a virtual obstacle course of spheres inserted into the ground. "Something you can trip on," Kidziński says. The purpose was to make the models more generalizable and to see if the model can adapt just as a human would. Those who join the competition create their controllers in OpenSim, which Kidziński and his colleagues modified to include reinforcement learning in a changeable environment.

More than 400 people have joined the competition, which received over 1,500 submissions by early October—three weeks ahead of the deadline for the competition's first round. For November, the top models from the first round have been given a new challenge—a change in muscle strength, for example, or the frequency of obstacles. The victor's model will be the one that goes farthest in a set amount of time.

While there are some cool prizes, most participants are in it for the learning and community recognition. The competition's leaderboard shows a thumbnail video of the models in action. In early October, the top contestant was **Jackie Tseng**, a PhD candidate at Tunghai University in China. She says the most interesting part of the competition has been training her agents and watching them progress from falling down continually, to taking first steps, to crossing over an obstacle, and then to running—and earning a high score. Another competitor, **Anton Pechenko**, a research and development engineer at Yandex, in Moscow, agrees this is the fun part: "There are three things you can watch forever," he says, "Fire, water, and your agent performing actions solving the problem."

Tseng looks forward to the day when computer vision might be added to reinforcement learning systems such as this one, to allow agents to understand and adapt to more complex environments. For now, she says, "Though the half-humanoid model trained in the 'Learning to Run' challenge is still relatively simple, it is a quite important beginning in AI for biomechanics." □

DETAILS

Results of the NIPS 2017: "Learning to Run" challenge will be posted at https://www.crowdai.org/challenges/nips-2017-learning-to-run

Acknowledgements:

Coordinators: Carmichael Ong, Mohanty Sharada, Jason Fries, Jennifer Hicks, Joy Ku

Advisors: Sergey Levine, Marcel Salathé, Scott Delp

The Mobilize Center for Mobility Data Integration and Insight is an NIH Big Data to Knowledge (BD2K) Center of Excellence at Stanford University.

BY KATHARINE MILLER

IS CLINICAL GENOMICS TESTING WORTH IT?

Cost-effectiveness studies yield answers to the complex question of whether clinical genomics testing has value.

Whole-genome testing has now reached the long-anticipated "$1,000 genome" level; and more targeted genetic panels cost even less. But the costs associated with genomic testing don't end with sequencing. Additional expenditures—for follow-up testing or treatments—may far exceed the investment in sequencing itself.

"I hear people say, 'of course it's cheaper and better to just sequence people up front: More information is better,'" says **Kathryn Phillips, PhD**, professor of clinical pharmacy at the University of California, San Francisco. "In fact, it might be better in some situations but not others." Phillips and others are trying to pinpoint the situations for which the health benefits of genomic testing outweigh the costs, using cost-effectiveness analyses.

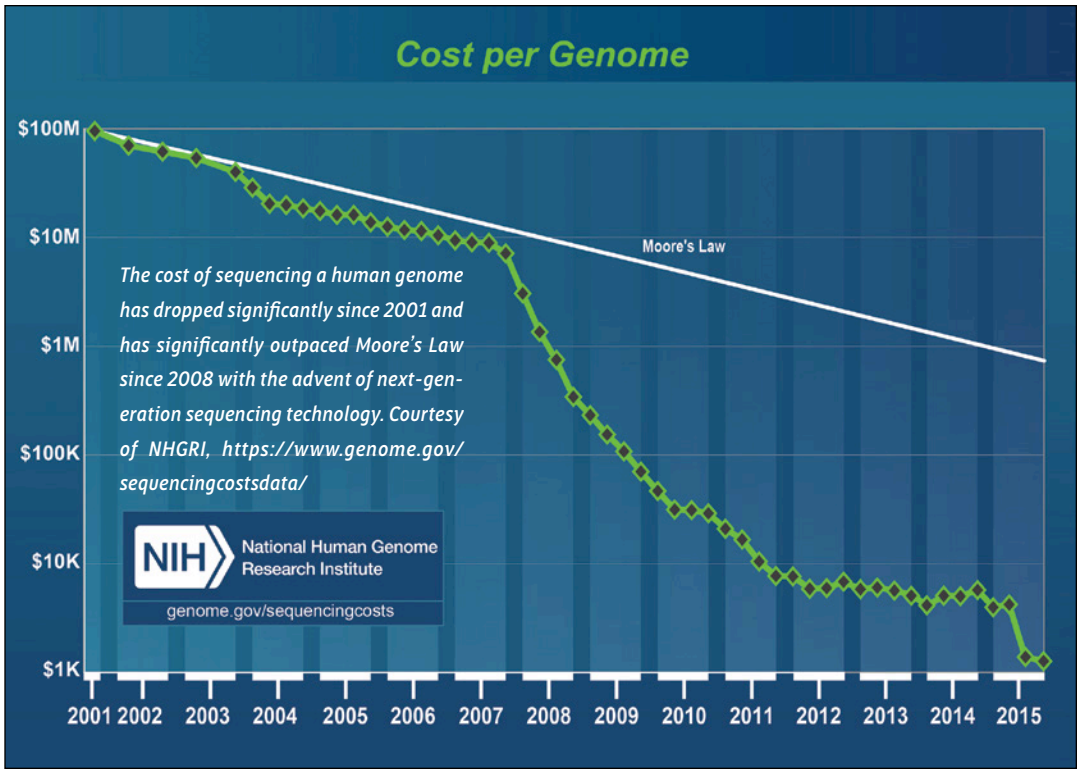
It's not a simple task: The inputs—such as the risks associated with a genetic variant and the possible benefits of testing—are uncertain. Sequencing also provides information about many different genes, and each variant will have a different cost-benefit ratio. "You can't do a holistic view of the full benefit of these tests," says **Eman Biltaji, PhD**, graduate research assistant at the University of Utah. "You can only do a study focusing on one piece of it and then another study focused on another piece of it." Finally, patients' preferences and behaviors complicate things. For example, if a patient who gets a negative genetic test decides to forego routine disease screening as a result, that could be a hidden cost of testing.

As payers (insurance companies and governmental insurers) weigh whether to cover the costs of testing or not, cost-effectiveness research may help provide some answers to the key question: how to deliver the right service to the right patient at the right time. Cost-effectiveness analyses are revealing valuable benefits for certain patients, in the areas of rare pediatric disease, cancer, and pharmacogenomics. But the jury is still out as to whether whole exome or whole genome sequencing (WES or WGS) for healthy patients is worth it.

Rare Diseases in Children: Test Early!

Children affected by rare monogenic conditions often undergo an extended diagnostic odyssey during which they are poked, prodded, tested and hospitalized at great expense to their families and the healthcare system. A study of 40 such patients published in *Genetics in Medicine* in January 2017 found that cost-effectiveness was maximized when patients were offered WES as soon as a problem was suspected. "If you find out what's happening early in the diagnostic trajectory, it does allow you to influence management of the genetic disorder a lot more than if you're provided with an answer a few years down the track," says **Zornitza Stark, MD**, a clinical geneticist at Murdoch Children's Research Institute in Melbourne, Australia, one of the lead researchers on the study.

Another similar study of 150 pediatric neurology patients in the Netherlands also found early use of WES to be cost-effective; and a Canadian study of 103 pediatric patients with suspected genetic disorders found that WGS provided a



far higher diagnostic yield (41 percent) than a gene panel (24 percent) or WES (34 percent), suggesting that WGS may be cost-effective in this context as well.

Studies like these have an impact because they can convince payers that genomic testing is worth reimbursing, Stark says. "Our state government has recently announced big money—$8 million—for rare disease diagnosis so that we can provide this type of testing."

Stark is now interested in pushing for rapid testing in hopes of returning WES results in a few days rather than 4 to 6 months. "It costs a lot more but potentially allows better decisions in ICUs (intensive care units) which are expensive places anyway." A day in the ICU costs about $4,500, she says, while rapid testing costs about the same. But the costs still need to be studied to determine whether testing actually makes a difference. "I think in the rare disease space, there's been an assumption that it doesn't really matter: These children are just considered incurable, untreatable. You're just giving it a name," she says. "That is sometimes true, but you'd be surprised by how much of an impact we've had on our patients. It has certainly exceeded our expectations."

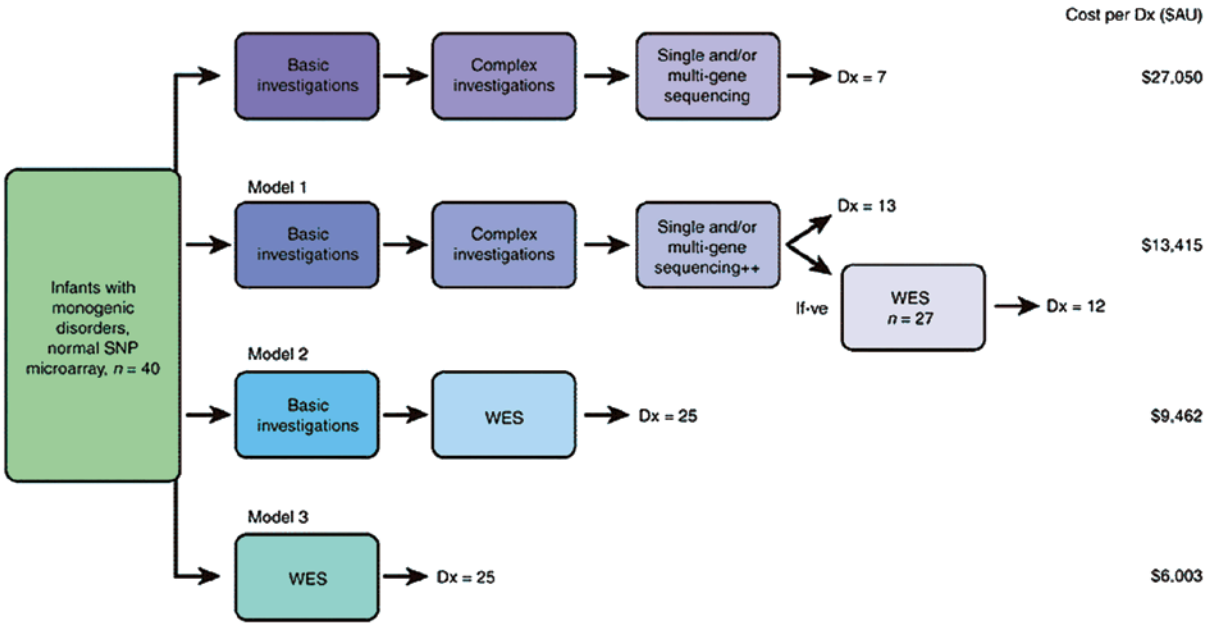
Cancer: Testing the Right Genes at the Right Time

In the cancer arena, some gene panels assess patients' risk of various cancers and others evaluate the genetic makeup of a specific tumor to determine the most effective treatment.

"In the cancer risk space, the big question is this: 'How big should the panels be?'" says **David Veenstra, PhD**, professor and associate director of the Pharmaceutical Outcomes Research and Policy Program at the University of Washington School of Pharmacy. "With more and more genes on the panels, payers become concerned that there might be things on there that don't have strong evidence behind them—i.e., that aren't that pathogenic."

In one study of colorectal cancer, he and his colleagues found not only that including highly penetrant colorectal cancer risk genes on a panel was cost effective, but also that including less penetrant genes added minimal additional cost and was therefore still cost-effective. They also found that testing relatives of colorectal cancer patients with highly penetrant pathogenic variants was cost-effective for the relatives in quality-adjusted life years—a standard measure of health benefit.

In the cancer treatment arena, the concerns are somewhat different. Finding the most effective treatment from the get-go can make a huge difference to patients; and next-generation sequencing panels can help guide chemotherapy treatment decisions. For example, about 40 percent of colorectal cancer patients have mutations in the RAS gene and therefore do not respond to certain adjuvant chemotherapies that work well for those without a RAS mutation. "Doing a test up front means that you protect patients from unnecessary, harmful, expensive treatments," Biltaji says. "It's much better to use



Stark and her colleagues compared the diagnostic trajectory and resulting diagnostic yield and costs per patient for 40 infants with monogenic disorders under four conditions including standard care (yield: 7 diagnoses) and three other models: (1) WES as a last resort after exhausting all standard investigations, including planned gene tests (yield: 13 diagnoses); (2) WES replacing some investigations, particularly gene sequencing tests, complex biochemical tests, and invasive tests, (yield: 25 diagnoses); and (3) WES replacing most investigations (yield: 25 total diagnoses). Model 3 was the most cost effective per diagnosis ($6003) while standard care was the least cost effective ($27,050). Reprinted by permission from Macmillan Publishers Ltd from Stark Z, Schofield D, et al., Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement, Genet Med *19: 867-874 (2017).*

these treatments for the right patients."

But is it best to do a single gene test for the RAS mutation or a gene panel that provides information about other genes implicated in treatment response as well? Another challenge: Testing a tumor's gene expression once isn't enough because the tumor evolves, Biltaji notes. If it is difficult to get new tumor biopsies—from the brain, for example—a patient might continue to be treated based on old, incorrect information. That would change the cost-effectiveness analysis.

Pharmacogenomics Panels: A Few Base Hits

Pharmacogenomic screening can reveal how a person's genes affect his or her response to drugs, leading to safer prescribing and dosing. But the question remains whether routine pharmacogenomics screening is cost effective.

In a recent study of elderly patients taking three or more medications, Biltaji and her colleagues found that, compared with matched controls, screened patients had a lower rate of hospitalizations and ER visits—but higher outpatient visits—during the ensuing four-month period. Overall, there was a cost saving in the genetic testing arm—but the dollar amount was small—$218 net savings per patient, including the cost of the test.

"In the pharmacogenomics space, people have been looking for a home run example where we'll be saving lives left and right and revolutionize medicine," Veenstra says. "But it's not about that; it's about a bunch of base hits." In order to justify pharmacogenomic screening of healthy people generally, he says, "Panel testing will need to be in the hundreds of dollars …. And I think that's where we are going."

Genome Testing of Healthy Folks: The Big Question Mark

Genetic panels and WES for people who are already ill is one thing. Whole genome sequencing or gene panels for otherwise healthy patients is another. "Usually tests are done for a particular reason," Phillips says. "If you're just fishing, then it's a question of how to put a value on that." There's also the question of whether integrating genome tests into healthy patient treatment could lead to overuse—or possibly even underuse—of healthcare resources.

"How do people who are 'negative' for risk genes behave?" wonders Veenstra. "Are they less likely to get a mammogram even though a negative BRCA finding doesn't mean they are at lower risk for breast cancer? You want people to follow recommendations but not pursue health care consumption behaviors that aren't justified."

There's also the question of how to handle incidental findings—i.e., the discovery of genetic variants that were not the target of the genetic test's original goal. For example, if a person is tested for colorectal cancer treatment purposes, but the test reveals variants with other medical implications, should the patient be informed of these results? This issue rose to the fore a few years ago when the American College of Medical Genetics and Genomics (ACMG) recommended that clinical laboratories performing genomic testing should routinely report any incidental findings relevant to 56 (at the time—now 59) genes considered actionable and having a high probability of causing disease. The recommendation was later revised to state that patients may opt out of receiving the findings.

Prior to making the recommendation, the ACMG had not evaluated whether returning incidental findings to patients would be cost-effective. In work published in *Genetics in Medicine* in 2014, Veenstra took on that challenge. He and his colleagues applied some clever strategies to determine that the return of incidental findings could prove cost effective for some patient populations. "Our work to date has shown that the reporting of incidental findings could be worth doing," Veenstra says. "But we need better estimates of penetrance and how people behave when provided this information."

A group at Brigham and Women's Hospital in Boston is trying to find the answer to that very question. As part of a large research project called MedSeq, **Kurt Christensen, PhD**, instructor in genetics and medicine at Harvard Medical School, and his colleagues performed WGS on primary care patients at a cost of about $5,250 per patient. They then looked at whether the information gained led to higher healthcare expenditures in the ensuing six-month period. Compared with untested patients, those with WGS results incurred about $350 more in healthcare costs during that time (a difference that was not statistically significant), Christensen says. The MedSeq team will continue to follow these patients long term. "Six months out is too soon to see the kinds of cost-savings you might see as a result of correct dosing or avoiding adverse reactions to medicines, or detecting or preventing disease," he says.

But there are methodological challenges to this kind of cost-effectiveness research. "What are the right methods for capturing services linked to the genetic information so that we can distinguish what was ordered in response to sequencing as opposed to other conditions that might arise?" he asks.

Knowing Enough to Make Good Decisions

Ultimately, the value of genomic sequencing lies in how it affects clinical practice. Veenstra predicts that in five to ten years, even run-of-the-mill healthcare systems will be considering screening untested populations for the ACMG and pharmacogenomic genes. And cancer gene testing is well underway in many healthcare settings. If the cost of WES and WGS continue to drop, perhaps they will become routine as well.

"The exciting part," Veenstra says, "is that there's a good chance these tests have good economic value." □

BY MARCUS WOO

PROTEINS FOR EVERY OCCASION

Protein design ascends to new heights

Scientists are now able to design, in principle, almost any protein they want—a feat that was inconceivable just a few years ago. They are reengineering existing proteins found in nature, as well as constructing proteins from the ground up, atom by atom.

Custom-designed proteins could mean new and better vaccines, drugs, and other therapeutics; precisely designed biosensors; and catalysts capable of producing chemicals and pharmaceuticals in a more environmentally friendly manner.

Custom-designed proteins could mean new and better vaccines, drugs, and other therapeutics; precisely designed biosensors; and catalysts capable of producing chemicals and pharmaceuticals in a more environmentally friendly manner.

Designed proteins may help solve many of the world's biggest problems, says **David Baker, PhD**. As a professor of biochemistry at the University of Washington in Seattle, he's been a pioneer in developing computational methods to build proteins from scratch, a process called *de novo* protein design.

Thanks to an improved understanding of how proteins fold, as well as advances in computing and genomic technology, experts say the field is now at an inflection point, with progress developing faster than ever.

"I'd like to think we can do most everything," says **William DeGrado, PhD**, professor of pharmaceutical chemistry at the University of California, San Francisco. "And I'd like to think we can do a lot more than nature can do."

Designing a Protein

Scientists have understood the basic principles behind protein folding since the 1960s: Electrostatic forces between and among the amino acids in a protein sequence pinch the chain, folding it into its lowest energy state—a flexible 3-D structure that changes in response to other nearby molecules. Since then, progress toward understanding how proteins reach their 3-D structures has been steady, including the first *de novo* computational design of a protein nearly 20 years ago and many other protein design successes since.

These days, researchers can fully model and create proteins from scratch using an advanced software package called Rosetta, developed by Baker's lab. Rosetta users start with a desired protein structure and allow the program to fill in the details. Specifically, users first define a desired backbone shape—the arrangement of alternating amino and carboxyl groups that are part of each amino acid and that link together to form a polypeptide chain. The computer then calculates how well various side chains (which differ for each amino acid) fit around that backbone to produce the desired structure. "If you put a side chain in one position, that can dictate what's on the neighboring position so they nestle together," says **Brian Kuhlman, PhD**, professor of biochemistry and biophysics at the University of North Carolina at Chapel Hill. Meanwhile, Rosetta ensures that the whole molecule is at its lowest, most stable, energy state.

Yet most of the sequences that Rosetta comes up with for a particular structure won't actually fold into a stable shape in the lab. And calculating backward to check whether the sequences do indeed generate the desired protein structure only gets you so far. To truly validate a sequence, one must synthesize the protein and test its stability. Until the advent of large-scale *de novo* approaches (see below), this required that proteins be designed and tested one at a time.

Neanderthal Design: Tweaking Natural Proteins

Most protein engineering to date has involved tweaking proteins found in nature to give them slightly different functions. Baker calls this Neanderthal protein design, similar to the strategy our primitive cousins would have employed—fashioning tools out of what was already lying around—for example, chipping away at a rock or sharpening a stick.

Baker's team published an exciting example of this strategy in *Nature Biotechnology* in June 2017: They designed a protein that prevents mice from getting the flu. They knew that the flu virus's surface contains a mushroom-shaped protein called hemagglutinin that enables the virus to infect cells by binding to a sugar molecule in the cell membrane. So they created a protein, dubbed flu glue, that can glom onto hemagglutinin, blocking it from infecting cells. It might not become medicine for humans anytime soon, but could

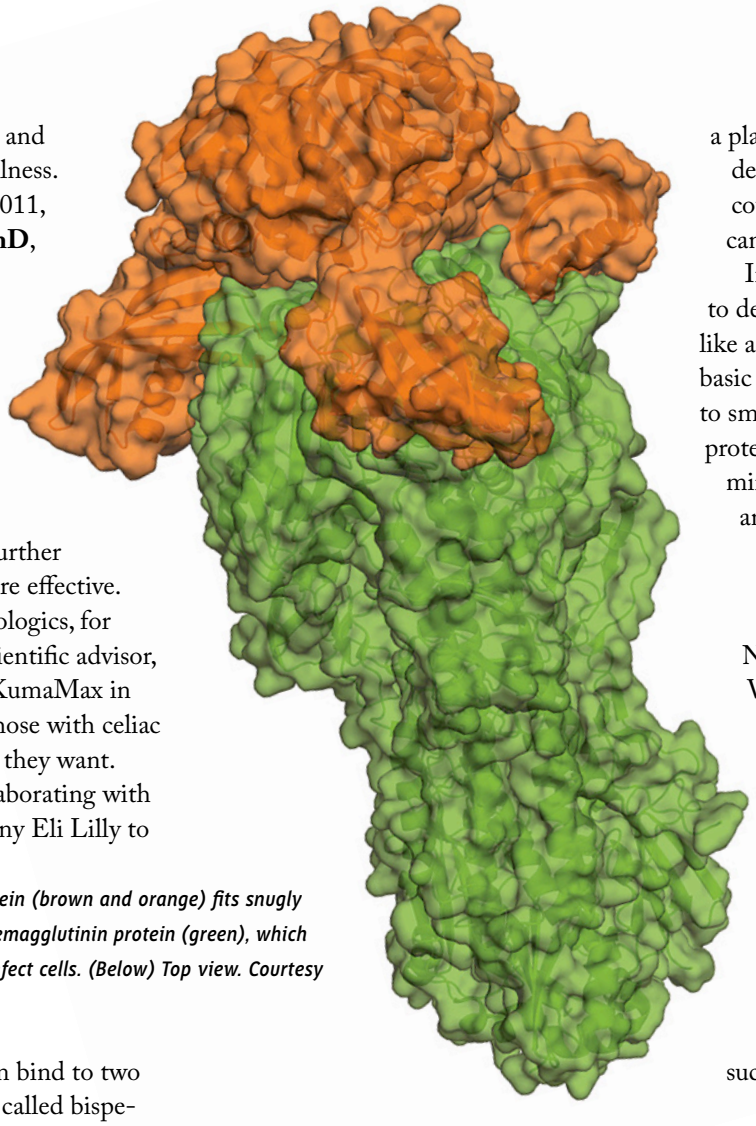
be used to develop a quick and easy way to diagnose the illness.

A few years earlier, in 2011, **Ingrid Swanson Pultz, PhD**, translational investigator at the University of Washington, led the development of an enzyme called KumaMax that breaks down gluten. Since then, the molecule's design has gone through further refinements to make it more effective. Pultz co-founded PVP Biologics, for which Baker serves as a scientific advisor, to further commercialize KumaMax in pill form. It might allow those with celiac disease to eat all the bread they want.

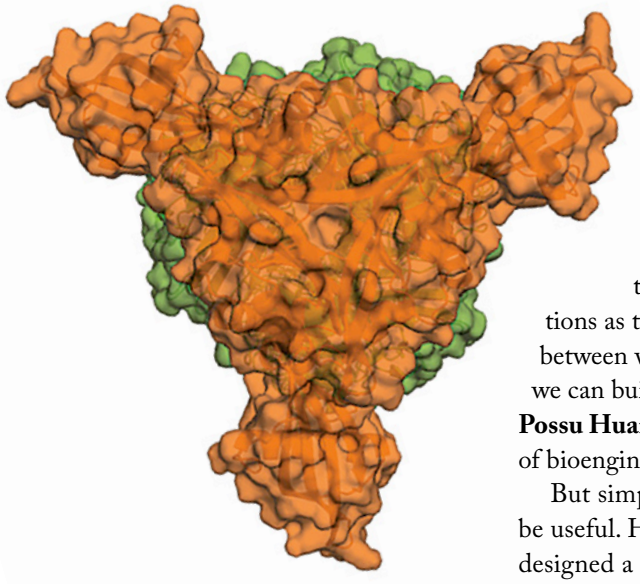
Kuhlman has been collaborating with the pharmaceutical company Eli Lilly to develop antibodies that can bind to two antigens at the same time, called bispecific antibodies. These kinds of antibodies can, for example, bind to both a tumor cell and an immune cell, thereby recruiting the body's immune system to help fight cancer. The trick is making sure they don't bind to other things in undesirable ways. In a 2016 paper published in *Structure*, Kuhlman's lab developed a strategy for predicting the specificity of bispecific antibodies.

By designing proteins that bind to specific molecules, researchers can also make new types of biosensors. In work published in *eLife* in 2017, for example, Baker's lab designed one that can signal the detection of the painkiller fentanyl. To test the sensor, the researchers incorporated it into a plant so the leaves turn color when it detects the molecule in question. This could ultimately lead to plants that can sense dangerous compounds.

In the future, Baker also wants to design proteins that can function like a rudimentary computer that does basic logic operations. This could lead to smart therapeutics such as designer proteins that can bind to a cell, determine whether it's healthy or sick, and release or not release a drug.



Flu glue: (Right) A designer protein (brown and orange) fits snugly on top of the influenza virus's hemagglutinin protein (green), which helps the virus latch onto and infect cells. (Below) Top view. Courtesy of Eva-Maria Strauch.



De Novo Design of Simple Proteins

Neanderthal design has its limits: When you start with protein backbones that were created through the evolutionary process, you miss out on a huge variety of options that nature never tried. By contrast, *de novo* design can explore the entire realm of possible protein backbones, some of which might have greater potential to prevent or treat disease than natural (or Neanderthal-designed) molecules such as antibodies or antibiotics.

For the most part, *de novo* efforts have been restricted to simpler proteins because more complex structures are beyond current computational capabilities. And researchers are still far from being able to design proteins with the same sophisticated functions as those in nature. "The gap between what nature can do and what we can build is still very wide," says **Possu Huang, PhD**, assistant professor of bioengineering at Stanford University.

But simpler proteins can still be useful. Huang, for example, has designed a donut-shaped protein called

a TIM barrel, which has potential as a biosensor and as a building block to construct larger molecules. And Baker's group has designed smaller proteins that can fit together to form so-called nanocages, which can serve as containers to deliver drugs.

Large Scale *De Novo* Design

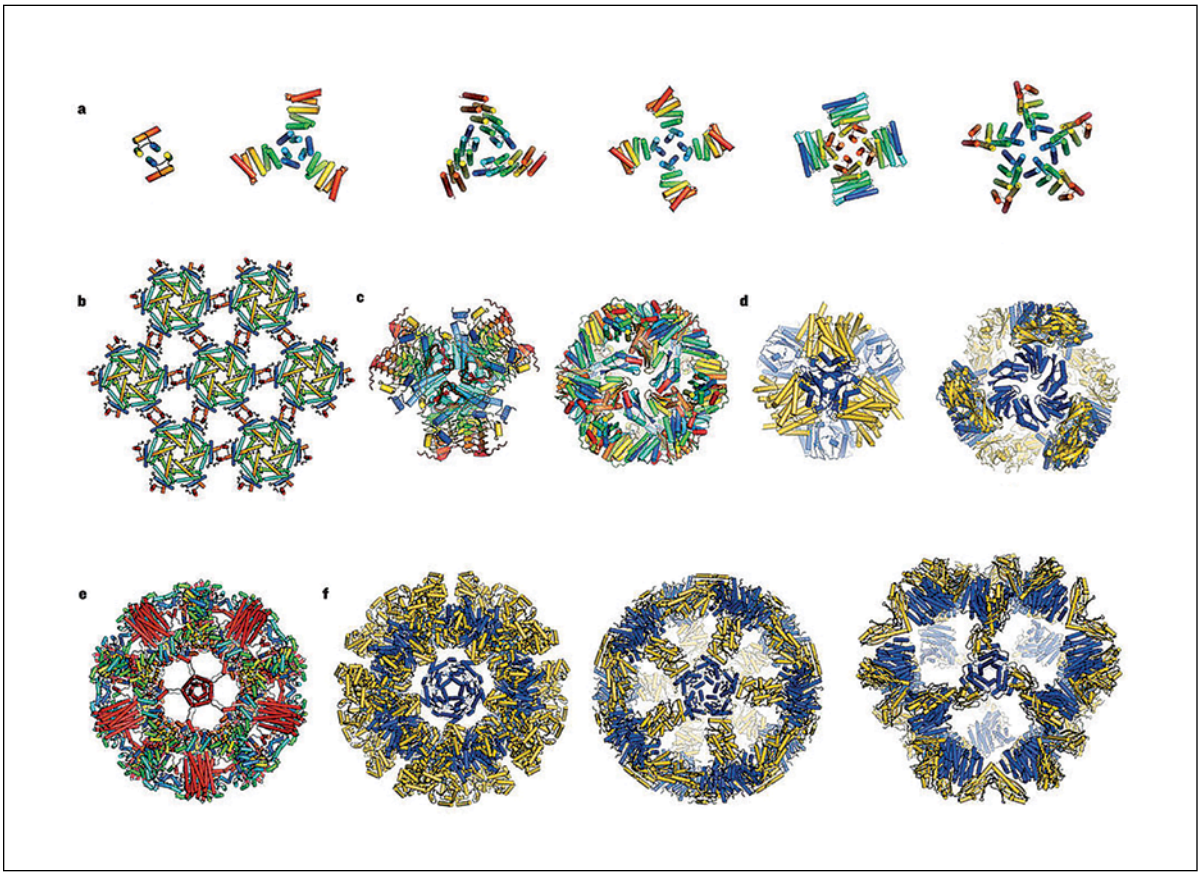
The process of generating *de novo* protein designs took a huge leap in 2017 with a pair of papers (one in *Science* and one in *Nature*) out of the Baker lab. They are noteworthy because they signal a new era of large-scale, data-driven *de novo* design.

In the July 2017 *Science* paper, the Baker group used Rosetta to design 15,000 mini-proteins. They then tested these proteins for stability using a novel high-throughput experimental approach. They engineered yeast cells that could encode the test proteins and ferry them to the cells' outer surfaces; and they bathed these cells in protease, an enzyme that breaks down proteins, to measure how the proteins reacted—to determine if they folded into a stable shape. At first, very few did. But the team analyzed the winners and losers to discover new rules of protein folding and incorporate winning features into the Rosetta pipeline. By the time they were done, their success rate for designing stable proteins had risen from 6 to 47 percent, and they had designed 2,788 novel proteins.

Such large-scale methods enable researchers to more efficiently design the proteins they want. "It gives you a lot of shots on goal," DeGrado says. But the effectiveness of the feedback loop is equally important and suggests that machine learning could be used to better harness the lessons learned from large-scale testing—an approach Baker is now pursuing.

In the October 2017 *Nature* paper, the Baker team added a new application to the large-scale strategy: drug design. They used *de novo* techniques to design and test 20,000 mini-protein drug candidates for targeting viruses (such as flu) or toxins. When tested, their designs for flu successfully protected mice from infection. Unlike the Neanderthal-designed flu glue, these designs bind on the side of the hemagluttenin protein to block the virus from fusing with a cell. Moreover, the mini-binders are small, easy to make, more stable than antibodies, and don't elicit much of an immune response (at least compared with modified natural proteins), suggesting that they may have potential as an anti-flu therapeutic.

Baker is optimistic. Neanderthal design is indeed useful today. But *de novo* design is still likely the future, he says. His lab has spawned a community of hundreds of protein designers, including Huang and Kuhlman. "There's just a lot of energy and momentum in this area now, with lots and lots of smart young people going into it," Baker says. "The future is very bright." □



Protein interface design methods have been used to create self-assembling, cyclic homo-oligomers (a), two-dimensional hexagonal lattices (b), various self-assembling cages (c-f) and one- or two-component assemblies with icosohedral symmetry and 60 (e) or 120 (f) subunits. Due to these de novo *assemblies' symmetry, they could be used for the presentation of antigens in vaccine applications as densely clustered antigens often lead to better immune responses. Also, the interior volumes of these cages are big enough to package cargo for delivery to targets. Reprinted with permission from MacMillan Publishers Ltd: Huang PS, Boyken SE, Baker D, The coming of age of* de novo *protein design,* Nature *537, 320-327 (15 September 2016).*

BY KATHARINE MILLER

BIG DATA FEEDS UNDERSTANDING OF OBESITY

How data about genes, physical activity, and the environment are yielding insights about obesity

Despite extensive efforts, a clear understanding of the obesity epidemic remains elusive. Scientists have implicated specific causal genes (40 to 60 percent of obesity is considered heritable); epigenetic effects (environmental changes to genes); and the microbiome (though this is controversial). Behavioral causes such as diet and exercise clearly matter, but behavior is hard to change; many people lose weight only to regain it, or begin an exercise regimen only to quit.

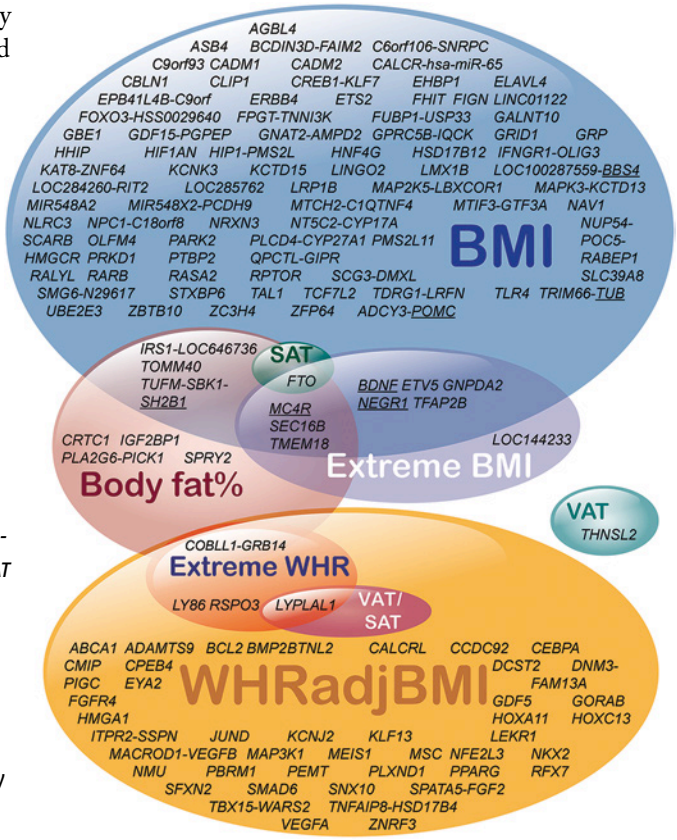
With so many variables at play and no one-size-fits-all solution, some researchers are turning to big data. "Obesity lends itself well to big-data approaches that can bring out the real signals despite noise and heterogeneity," says **Elizabeth Speliotes, MD, PhD, MPH**, associate professor of internal medicine and computational medicine & bioinformatics, University of Michigan Medical School.

Big data is now beginning to cut through some of obesity's opacity, including clarifying roles played by genetics, physical activity, and the environment. But relating all of the different and complex variables will require more data and, in particular, more integrated data.

Obesity and Genes

Body mass index (BMI—weight over height squared) is considered a useful way to measure overall obesity, while waist-to-hip ratio (WHR) captures obesity that is centered in the abdomen. In two papers published simultaneously in *Nature* in 2015, Speliotes and her colleagues conducted genome-wide association meta-analyses of BMI in nearly 340,000 people, and of WHR (adjusted for BMI) in nearly 225,000 individuals. They found that a different set of gene variants predispose a person to a high BMI compared to those associated with a high WHR. The genes linked to high BMI were largely based in the nervous system whereas those linked to high WHR related to adipose differentiation—i.e., the development of fatty tissue. High WHR is also more closely linked to health-related complications such as diabetes and cardiovascular disease. "There may be a more protective effect of depositing the fat subcutaneously rather than in the belly," Speliotes says. Researchers are currently exploring what the discovered genes do, how they do it, and how they are connected to each other.

Her team is also finding that the complications of obesity don't always follow from a predisposition to obesity. Other genes may sometimes intervene to provide either a protective or detrimental effect. With more and bigger datasets will come a better sense of how these interactions work, Speliotes says. "The combination of genes will be more and more predictive."



Different genetic loci are associated with different measures of obesity, as shown in this Venn diagram, which identifies genetic loci associated with BMI, body fat percentage, waist-hip ratio adjusted for BMI (WHRadjBMI), visceral adipose tissue (VAT), subcutaneous adipose tissue (SAT), the VAT/SAT ratio and extremes of BMI and WHR. Reprinted with permission from Fall T, Mendelson M, Speliotes E, Recent Advances in Human Genetics and Epigenetics of Adiposity: Pathway to Precision Medicine?* Gastroenterology *152:7:1695-1706 (May 2017).

Obesity and Physical Activity

Historically, most studies of physical activity and obesity were relatively small; relied on individuals' self-reports, which aren't entirely reliable; and reported the respondents' average activity levels—resulting in a further loss of information.

Wearable activity monitors have changed all that. Sample sizes have blossomed and detailed information is abundant. Researchers at Stanford University's Mobilize Center recently analyzed 68 million days of physical activity data for 717,527 people in 111 countries collected from a smartphone fitness application (Azumio's). The study, published in *Nature* in 2017, found that the disparity of physical activity distribution within a country, which they dubbed "activity inequality," is a better predictor of a country's obesity prevalence than average activity levels.

"The activity inequality concept really

jumped out at us," says **Tim Althoff**, a PhD candidate in computer science at Stanford University. When the team plotted individuals' steps per day in each location or country, a routine calculation done just to look for outliers, he says, "some of these curves were much wider than others." In Saudi Arabia, for example, there are fewer people on the low end of the norm, while in Japan, the curves are narrower with relatively few people walking way more or way less per day than typical. And the biggest revelation: "We found that in countries that are more unequal in activity, activity in females is reduced disproportionately." When they modeled possible interventions, it turned out that "if you just lifted women's activity levels to the level of their male counterparts, you would cut activity inequality by half," Althoff says.

Activity data can also be connected to genetics. For example, **Misa Graff, PhD**, research assistant professor of epidemiology at the University of North Carolina Gillings School of Global Public Health, recently did a meta-analysis of results from 60 genome-wide association studies (GWAS) covering more than 200,000 adults, looking at the interaction between obesity genes and physical activity. Although her data for physical activity was the old-fashioned self-reported kind, she did identify genes that are influenced by physical activity and, by controlling for physical activity, new obesity-related genes.

Still, she says she's "not very satisfied, actually.... We could find more if we had better measures of physical activity." Graff would love to use data from smartphones to look at how active people really are and coordinate that with eating and sleeping habits. "Without data that is harmonized across individuals, you can't get at a lot of the questions about genes," she says.

Figure 1

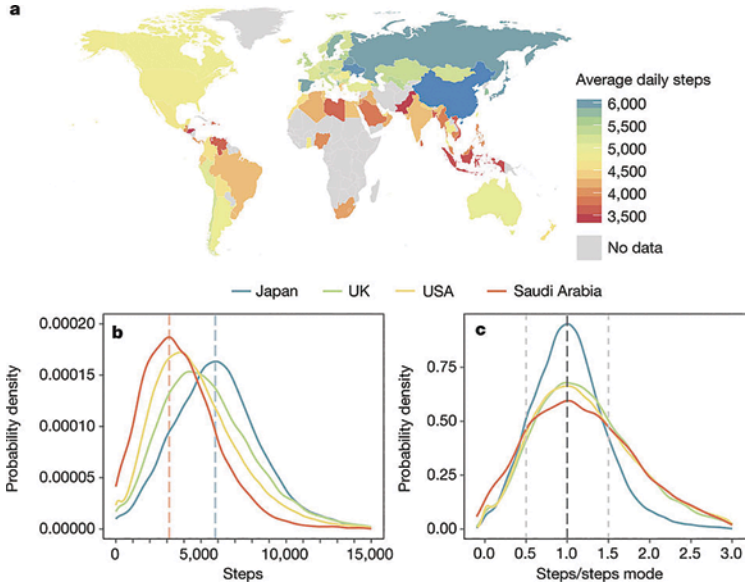
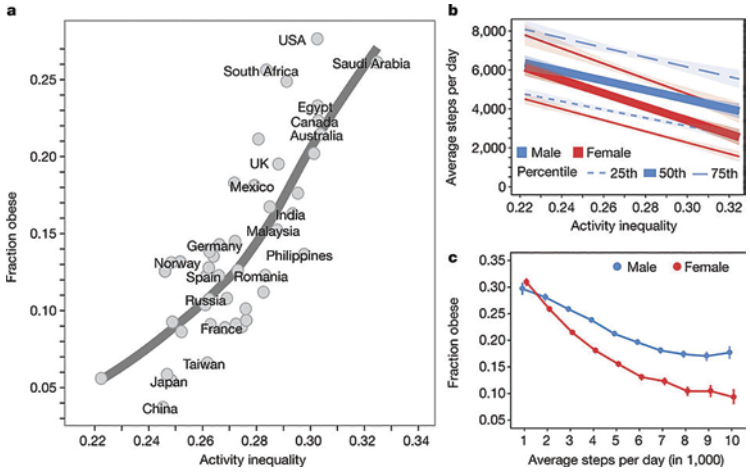


Figure 2



Obesity and the Environment

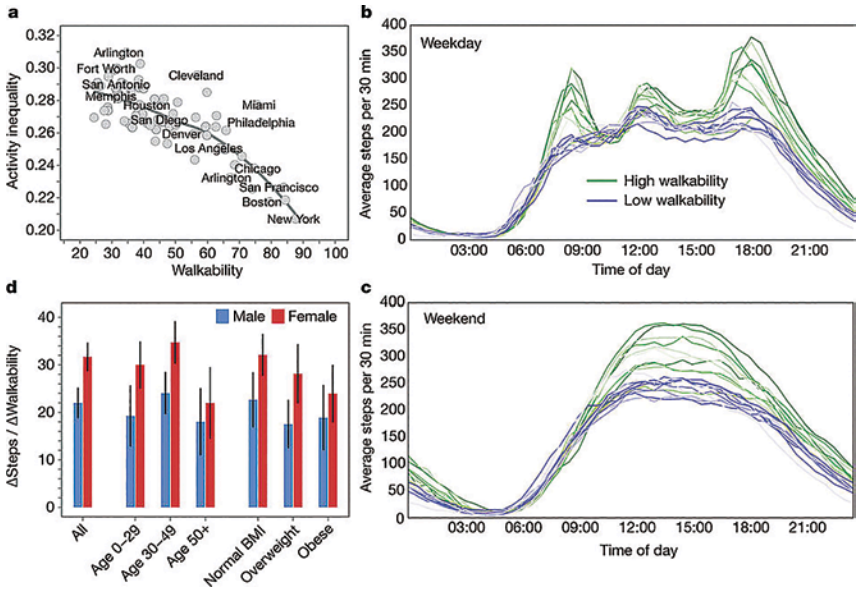
Big data could be useful in understanding obesity-environment connections at multiple scales: the epigenetic level (the ways that genes are turned on and off in response to environmental influences); the microbiome level; and even at the urban infrastructure level.

When it comes to how the microbiome and the epigenome affect obesity, the

Stanford's Mobilize Center researchers analyzed smartphone data from more than 68 million days of activity by more than 700,000 users of the Azumio fitness app. They discovered significant variability in mean daily steps by users in 111 countries across the world (figure 1a). Moreover, the distribution of steps varied by country, as shown in 1b and 1c for four representative countries. This observation led the researchers to define the concept of "activity inequality," which proved to be a useful predictor of obesity rates (figure 2a), especially among women (figure 2b and 2c). For both males (blue) and females (red), a larger number of steps recorded is associated with lower obesity, but for females, the prevalence of obesity increases more rapidly as step number decreases. Reprinted by permission from Macmillan Publishers Ltd: Althoff T, Sosič R, Hicks JL, King AC, Delp SL, & Leskovec J, Large-scale physical activity data reveal worldwide activity inequality, Nature *547, 336-339 (2017).*

jury is still out. In a recent meta-analysis of the gut microbiome, researchers at the University of Michigan found only a weak association between the microbial communities found in human feces and obesity status. And although epigenetic changes might influence obesity risk, the specifics are still unclear, Speliotes says. Moreover, she notes, obesity itself might have consequences for epigenetics. Indeed, in a 2017 *Nature* paper, an epigenome-wide association study of over 10,000 people found that BMI is associated with widespread changes in DNA methylation (a common epigenome marker), and that these changes predict future development of type 2 diabetes.

On the infrastructure front, however, the evidence is clearer: Certain neighborhood attributes—such as high poverty and crime, a lack of physical activity amenities, and a lack of transportation and recreation infrastructure—are associated with lower physical activity and higher obesity. But Graff found that, in a cohort of nearly 8,000 adolescents, physical activity helped reduce BMI regardless of the neighborhood-level factors. "Genes influence the BMI at the given activity level regardless of where you live," Graff says. While this result might seem unsurprising, it reinforces the value of higher levels of physical activity in obesogenic neighborhoods—a policy promoted by



Althoff and his colleagues found that higher walkability scores (from WalkScore.com) for 69 US cities are associated with lower activity inequality (a). Highly walkable cities show greater spikes in weekday walking during commute and lunch hours (b) as well as on weekends during daytime hours (c). Higher walkability is also associated with more daily steps across age, gender and BMI groups (d). Bars show the steps gained per day for each point increase in walkability score for 24 cities in the US. Reprinted by permission from Macmillan Publishers Ltd: Althoff T, Sosič R, Hicks JL, King AC, Delp SL, & Leskovec J,, Large-scale physical activity data reveal worldwide activity inequality, Nature *547, 336-339 (2017).*

whether or not a walkable city actually causes higher physical activity levels, only that the two are related. To get at the causality question, Althoff will look at people in his dataset who move between cities to see if activity and obesity status change in response to the walkability of the new location. In future work, Althoff also wants to look at food and nutrition patterns using data from people who track that on their devices; as well as how weather or changes to a transportation network affect activity levels. "Sensor-equipped smartphones and watches allow us to get into more detail regarding how physical activity changes with all kinds of factors," he says.

The Big Data Difference

One-third of the world is now obese; consumer wearables are collecting activity and diet information globally; and genetic and genomic datasets are exploding as well. Can all of these sources of big data converge to help us make sense of the various causes and potential solutions to this complex condition called obesity? It remains to be seen, but these data sources are certainly feeding an insatiable appetite for understanding. ☐

One-third of the world is now obese; consumer wearables are collecting activity and diet information globally; and genetic and genomic datasets are exploding as well. Can all of these sources of big data converge to help us make sense of the various causes and potential solutions to this complex condition called obesity?

former first lady Michelle Obama.

Althoff also considered infrastructure in his *Nature* paper. He and his colleagues found that physical activity levels were higher (and activity inequality was lower) in more walkable cities. However, their analysis cannot reveal

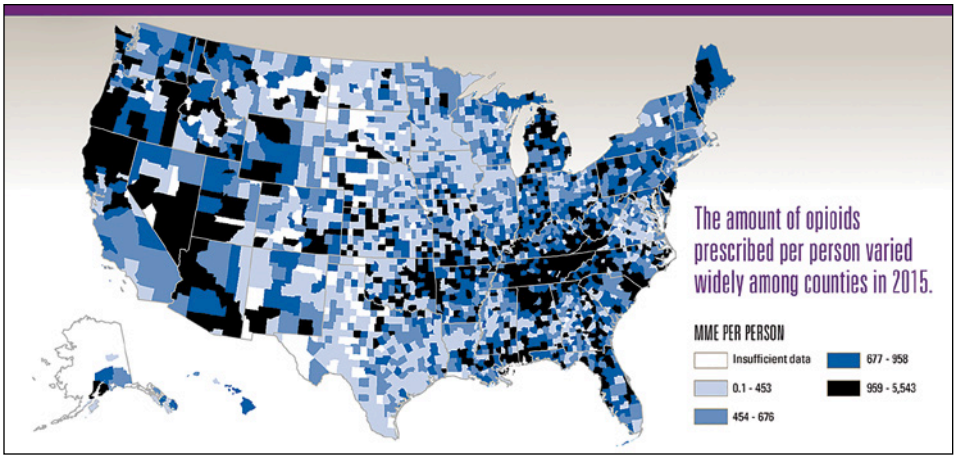
BY SARAH WILLIAMS

CAN BIG DATA STOP OPIOID ABUSE?

Armed with electronic health records and insurance claim information, data scientists are trying to predict who's going to become addicted to opioids—and stop them before it's too late.

A person who is overdosing on opioids exhibits telltale signs: a limp body, slowed breathing and heart rate, and blueish or purplish fingernails and lips. But millions more who live with an opioid problem are harder to spot. They may be male or female, old or young, employed or homeless, lonely or part of a large supportive family. It's often an invisible problem until it's too late.

To identify those who are abusing opioids as well as those who are at risk, researchers are turning to datasets—including electronic health records (EHRs), insurance claims, statewide pharmacy databases, and emergency medical service calls.



Higher opioid prescribing puts patients at risk for addiction and overdose. The wide variation among counties' MME (morphine milligram equlvalents) per person (as shown on this map) suggests a lack of consistency among providers when prescribing opioids. SOURCE: Centers for Disease Control, Vital Signs, *July 2017.*

"The more knowledge we have and the more we can use automated tools to fight this, the better," says **Caleb Alexander, MD**, co-director of the Johns Hopkins University Center for Drug Safety and Effectiveness.

Indeed, data science could become an important weapon in the battle against the skyrocketing rates of opioid abuse in the United States. Though some opioids are street drugs, such as heroin, many people become addicted after being prescribed pain relievers that are considered safe in small doses and for a short period of time. But these drugs also cause a sense of euphoria that can lead people to crave them. As the cravings increase, prescription users may end up taking opioids without a prescription or in a larger dose than prescribed. The epidemic has been linked not only to increasing deaths by overdose (according to the Centers for Disease Control, in the US roughly 100 people die of an opioid overdose each day), but also to staggering medical costs, a rise in the number of children entering foster care, and even a shrinking labor force.

It can be difficult for clinicians to identify addicts or other individuals at high risk of abusing opioids. Even armed with a thorough physical and mental health history, the most skilled doctor may miss the signs; and that's where hard data comes in. Researchers are finding that information on when and where patients fill prescriptions, what doctors they visit, and where they live, among other factors, may reveal an addiction or risk of overdose. Using data to flag these individuals may be one part of a much broader effort to slow the opioid epidemic.

Calculating Risk from Claims and EHR Data

Thomas Ciesielski, MD, a doctor of internal medicine at Washington University School of Medicine, recently collaborated with Express Scripts, a pharmacy benefit manager, to determine whether they could predict opioid abuse or dependence using pharmacy and health insurance claims information. He and his colleagues analyzed de-identified data on nearly 700,000 Express Scripts users and discovered 12 patient characteristics that increased the odds of opioid abuse or dependence. The strongest predictors: chronic use of opioids, mental illness, alcohol and non-opioid substance abuse, younger relative age, and male gender. The work was published in the *American Journal of Medicine* in 2016.

"My hope in adding to this body of literature was to give clinicians a better understanding of the risk factors that a patient sitting in front of them has," says Ciesielski. "If they have this information before they write a prescription, they can have more informed discussions with their patients."

Joseph Boscarino, PhD, an epidemiologist and social psychologist at Geisinger Health in Danville, Pennsylvania, says that pinpointing a patient who may be addicted to opioids—based on EHR data—can be useful even after initial prescriptions are written. Boscarino and his colleagues recently discovered that a patient's healthcare costs often spike just before an overdose.

"We found the signal by accident," he says, pointing out that a spike in costs after an overdose—when a patient needs life-saving medical care and mental health support—would be more expected. When Boscarino and his team looked closer at their data—which included the electronic health records of more than 2,000 patients admitted to the hospital for an overdose between 2005 and 2015—they found other

risk factors associated with overdoses: being unmarried, unemployed, and taking other prescriptions along with opioids.

"Now our health system is aware that we might be able to catch overdoses by looking at this cost signal, especially if we know a patient is on all these other drugs," says Boscarino.

Moving Beyond the EHR

But while researchers like Ciesielski and Boscarino can find trends in their datasets, they also admit their limitations; the EHR or pharmacy claims datasets weren't created with the goal of diagnosing opioid misuse.

Ciesielski, for instance, suspected that the distance between a patient's home and where they had a prescription filled might be predictive of opioid misuse. "Patients driving further to get opioids might be doing it because they're no longer able to get opioids in their own area," he says. But the location data included in the Express Scripts dataset he studied only included zip codes—not the most precise way of calculating distance—and he found no clear association.

"Lots of information is just not in the EHR," points out Boscarino. Because many EHR systems are not connected to one another, "I don't know from the EHR whether a patient is also driving to another hospital system to get drugs," he says.

And studies are also plagued by the challenge of identifying opioid abusers in the first place—information that's needed in order to train a machine-learning algorithm. Today, most studies rely on International Classification of Diseases (ICD) diagnostic codes, which can be found in medical charts and claims data. But doctors, even if they suspect a patient may be misusing opioids, don't always put these codes for opioid abuse or dependence in the patient's chart. Instead, they may simply stop treating the patient or order a lower dose of opioids without documenting why.

Rather than relying on the ICD codes, **Brandon Cosley**, **PhD**, a researcher on the predictive analytics team at BlueCross BlueShield (BCBS) of Tennessee, says that they set a threshold for what is considered opioid abuse by looking directly at raw claims data in their possession.

"Our definition usually contains how much opioid has been prescribed, how many doctors someone has gotten prescriptions from, and how many pharmacies they've filled the prescriptions at," says Cosley. He parses patient claims data and demographics to find what other factors predict the elements of the definition. For several different subpopulations of BCBS Tennessee's members, Cosley says he has developed models that predict opioid abuse with 80 percent accuracy.

"We're talking about hundreds of different risk factors for any given individual, and they combine in many ways," he says. "The relative contribution of any one risk factor may be small."

Cosley has access to a more complete dataset than researchers studying single EHR systems, yet he says more data would be useful.

"I think one of our most exciting initiatives is to really incorporate some of the free-form text data that we get from our membership," Cosley says. "Things like nurses' and doctors' notes and customer service calls." He says BCBS Tennessee has technology that allows them to start analyzing this kind of data and is working on ways to use it effectively.

Other researchers are also eyeing whether genetics data might help predict opioid abuse. A recent research effort led by researchers at Proove Biosciences, a precision medicine company based in Irvine, California, that is dedicated to optimizing the treatment of chronic pain, studied an algorithm—dubbed the Proove Opioid Risk (POR)—which used small variations in the genome, called single nucleotide polymorphisms or SNPs, combined with clinical risk factors to predict opioid abuse. Patients in the highest category of POR scoring had 16 times greater odds of opioid use disorder. The work was published in 2017 in *Pharmacogenomics and Personalized Medicine*.

"I think in the future we're going to have to do a better job linking EHRs to other assets including genetics," Boscarino says. "Addiction is really complicated."

Data-driven Interventions

Some communities are also crunching data to identify doctors or pharmacies that are part of the problem, or neighborhoods that should be the focus of community-based interventions. Massachusetts, for instance, has used predictive analytics to allocate resources to neighborhoods with the biggest overdose rates. And Alleghany County, Pennsylvania, tracked overdose deaths in the county over a six-year timespan to pinpoint who was overdosing and when.

If communities, hospital systems and insurance companies can start recognizing the patterns of opioid abuse in their computers, can they stop the opioid epidemic? Probably not, but they might help slow it, says Alexander, who in July 2017 published a review in the *Journal of the American Medical Informatics Association* evaluating 15 algorithms that have been used to identify non-medical opioid use in EHR data.

"I don't think anyone is naive enough to believe that automated tools alone will suffice," he says. "But these tools can, in some cases, be used to simply raise awareness and promote information sharing across clinical teams when a patient is at elevated risk for injury or death."

The Geisinger pharmacy team, for instance, has started flagging opioid abuse risk factors in a patient's EHR, based on Boscarino's findings. This doesn't mean a patient with risk factors won't be able to get appropriate painkillers, but it could mean that a clinician or pharmacist reconsiders dosing or limits how many pills a patient gets at a time.

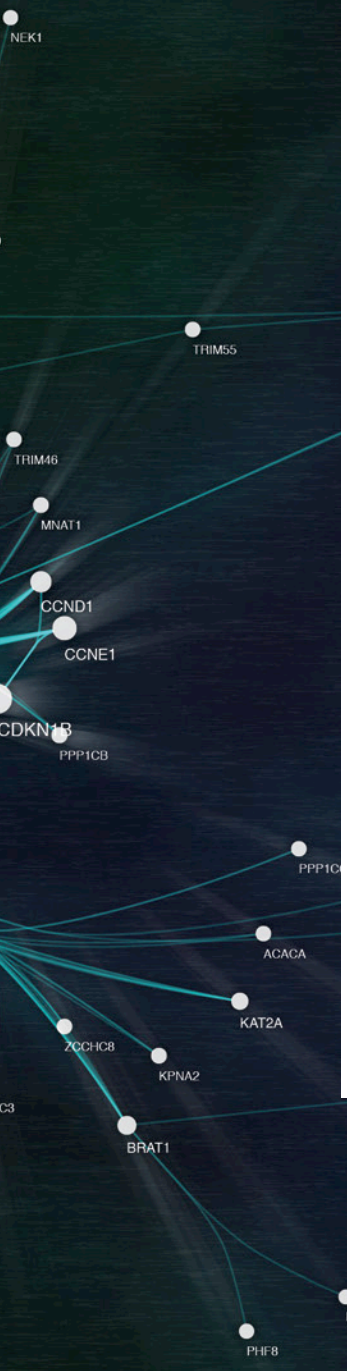
Alexander notes that insurance companies have a number of opportunities to improve patient care for those in pain while simultaneously reducing the overuse of prescription opioids. "Payers have a lot of tools in their toolbox, ranging from improving the coverage of non-drug treatments to designing new programs to identify and manage patients who are at highest risk of injury or overdose death," he says.

Data-science approaches to the opioid epidemic also have limitations: "They have to be used carefully and for the right purposes," Alexander says. "But they nevertheless can be quite powerful because they shine a light on potentially concerning patterns and allow for the identification of subpopulations who are at risk." □

BY KRISTIN SAINANI, PhD

NETWORK

Converging on Answers



To parents, the symptoms of autism can seem to appear from out of the blue during a child's first few years of life. But in recent years, researchers have shown that genes involved in the disorder likely affect neurodevelopment in the fetal brain. Among other implications, the results suggest that autism develops *in utero*, and is not due to exposures after birth.

To reach these conclusions, researchers needed more than the list of 65 strongly implicated autism-associated genes identified by genetics researchers. They needed an understanding of how the genes wire together into biological pathways that manifest as autistic traits—an understanding that emerged from an approach called network biology.

The story is similar for other complex diseases such as diabetes and Parkinson's disease: Lists of disease-associated genes are growing rapidly thanks to advances in sequencing technology, and network biologists are manipulating those lists to identify the larger biological pathways that malfunction in disease.

Having the gene list is only the first step, says **Trey Ideker, PhD**, professor of medicine and bioengineering at the University of California, San Diego. It's akin to having the parts list for an IKEA piece of furniture without the rest of the assembly manual. "Diseases involve networks of genes; and you have to map those networks if you're going to understand those diseases," says Ideker, whose network analysis tool, Cytoscape, has been cited more than 12,000 times. "What network biologists are trying to do is apply systematic approaches to map this wiring diagram."

Network biologists draw graphs that are essentially webs of biological relationships where the nodes are entities such as genes, proteins, or even patients, and the links between two nodes (called edges) represent specific interactions. For example, in a protein-protein interaction network, two proteins are connected by an edge if they are known to physically interact. By applying statistical and mathematical algorithms to these graphs, scientists are able to gain insights—such as identifying sets of genes that work more closely with each other than with other genes in the network, and thus may participate in the same biological pathway.

Network graphs often involve so many crisscrossing edges that they are referred to as hairballs—an indication that they can seem, to the uninitiated, nearly impossible to interpret. Add to this the fact that "interactomes" change in different cell types, tissues, and developmental phases, and you begin to get a picture of just how complicated this field is.

Network image supplied under Creative Commons License, created by Keiichiro Ono, UC San Diego.

BIOLOGY: to Complex Diseases

"In terms of network biology, we're approximately where genomics was in the late 1980s," Ideker says. Still, there's been considerable progress in recent years. Scientists have now mapped large-scale—if incomplete—networks for numerous organisms including humans; and network approaches have been at the heart of recent breakthroughs in complex diseases, including autism.

Stunningly, several autism research groups have used network biology to independently arrive at similar conclusions. "What's exciting is that multiple groups have used multiple different approaches to try and identify convergence among the set of genes that have been implicated in autism, and they're all coming up with very consistent findings," says **Jeremy Willsey**, **PhD**, assistant professor of psychiatry at the University of California, San Francisco.

Network biology encompasses a wide range of network types, including many based on physical interactions between and among cellular components (e.g., co-expression networks, genetic interaction networks, metabolic networks, protein-protein interaction networks, protein-DNA interaction networks, and protein-RNA interaction networks) as well as others based on similarity among patients or diseases. Each of these offers distinct biological clues that may help scientists transform their cellular parts list into insights about complex diseases. This article describes progress in a subset of these and looks at attempts to integrate the different types of networks.

Co-expression Networks: Genes Working Together in Autism

In a co-expression network, genes are linked if their expression levels are highly correlated. "If genes are truly operating in the same pathways we would expect them to be turned on or off at the same time," Willsey explains. Co-expression networks boast some of the earliest and most striking success stories in network biology, likely because the requisite data—microarray or mRNA-seq data—are relatively cheap to generate and readily available from existing studies.

In a 2013 paper in *Cell*, Willsey and colleagues leveraged co-expression networks to gain a foothold into the biology of autism. They identified nine high-confidence autism risk genes (greater than 97 percent chance of involvement) and 122 probable autism genes (greater than 50 percent chance of involvement). "We then asked the question: Is there a particular point in brain development and a particular region of the brain where the genes are most highly co-expressed, which may indicate that this is a relevant point in development for pathogenesis?" Willsey says.

The team obtained genome-wide expression data from 13 developmental stages and four groups of human brain regions from the BrainSpan database. They created 52 co-expression networks by linking each of the nine high-confidence autism genes to its top-20 co-expression partners (out of nearly 17,000 genes) for a given developmental stage and brain region. They reasoned that if a particular network was relevant to autism it should be enriched in probable autism genes as well. Indeed, unexpectedly high numbers of probable autism genes popped up in two networks, both involved in mid-fetal development. "We can see that this enrichment is very specific to particular time periods and brain regions, namely mid-fetal development in the prefrontal cortex," Willsey says.

They further localized the relevant co-expression to glutamatergic neurons residing within deep layers of the prefrontal cortex. In separate work, their group and others have also consulted the gene ontology for clues about biological function. "Every gene has a bunch of tags and we just looked for over-representation of those tags," Willsey says. Two key themes emerged: synapse biology and transcriptional regulation. "This was a very exciting moment for the field because we were able to go from genes to a specific hypothesis about pathogenesis," Willsey says.

In the same issue of *Cell*, a second, independent group of researchers—led by **Daniel Geschwind**, **MD**, **PhD**, professor of neurology and of psychiatry and biobehavioral sciences at the University of California, Los Angeles—reported strikingly similar results. "They saw similar convergence in mid-fetal development in the prefrontal cortex. They also observed enrichment in the same glutamatergic neurons," Willsey says.

Geschwind's team first built co-expression networks that were agnostic to autism risk genes: They used BrainSpan data to discover 12 "co-expression modules"—waves of highly co-expressed genes—that characterize normal fetal and infant brain development. They then looked for enrichment of suspected autism genes within these modules. Autism genes were over-represented in two modules involved in transcriptional regulation and three modules involved in synapse formation during fetal development. And the genes were highly expressed in glutamatergic neurons in the prefrontal cortex. A separate set of risk genes for intellectual disability were not enriched in any of the 12 co-expression modules, suggesting that autism is biologically distinct from intellectual disability.

"I think for the field it was very nice to see two different approaches come to a very similar conclusion in terms of pathogenesis," Willsey says. Other groups have since arrived at similar conclusions using network biology approaches. "For me that's particularly exciting, because historically in psychiatric disorders there has been a lack of agreement in the field about a lot of different aspects of biology," Willsey says.

Co-expression networks can also be used to implicate additional disease genes. "We can improve gene discovery by essentially using a guilt-by-association method," Willsey says. His collaborators at Carnegie Mellon University (**Kathryn Roeder**, **PhD**) and University of Pittsburgh (**Bernie Devlin**, **PhD**) developed an algorithm, Detecting Association with Networks (DAWN), that identifies hot spots within co-expression networks—areas where multiple autism risk genes cluster together. Genes that reside in these hot zones are automatically suspect, even if they've never been implicated before. "Genes that may not have had enough genetic evidence for association get their

scores strengthened if they're highly co-expressed with strongly associated genes," Willsey says. DAWN could be applied to other complex diseases as well, he says.

Genetic Interaction Networks: Interacting Double Mutants in Cancer and Parkinson's Disease

In a genetic interaction network, two genes are linked if a mutation in one alters the effect of a mutation in the other. For example, mutating a gene in either the BRCA DNA repair pathway (a pathway implicated in breast cancer) or the PARP DNA repair pathway alone is insufficient to kill the cell; but hitting both at once is lethal. The cancer drug olaparib exploits this so-called "synthetic lethality" by disabling the PARP pathway in cancer cells that already have a BRCA mutation. Mapping genetic interaction networks is costly and time consuming, so we can't yet approach genome-wide coverage. "Currently it's about 100 genes that are able to be interrogated by mere mortals," Ideker says. But genetic interaction networks offer more immediately actionable insights than co-expression networks, such as suggesting novel drug targets. So Ideker set out to make the process cheaper and less time consuming.

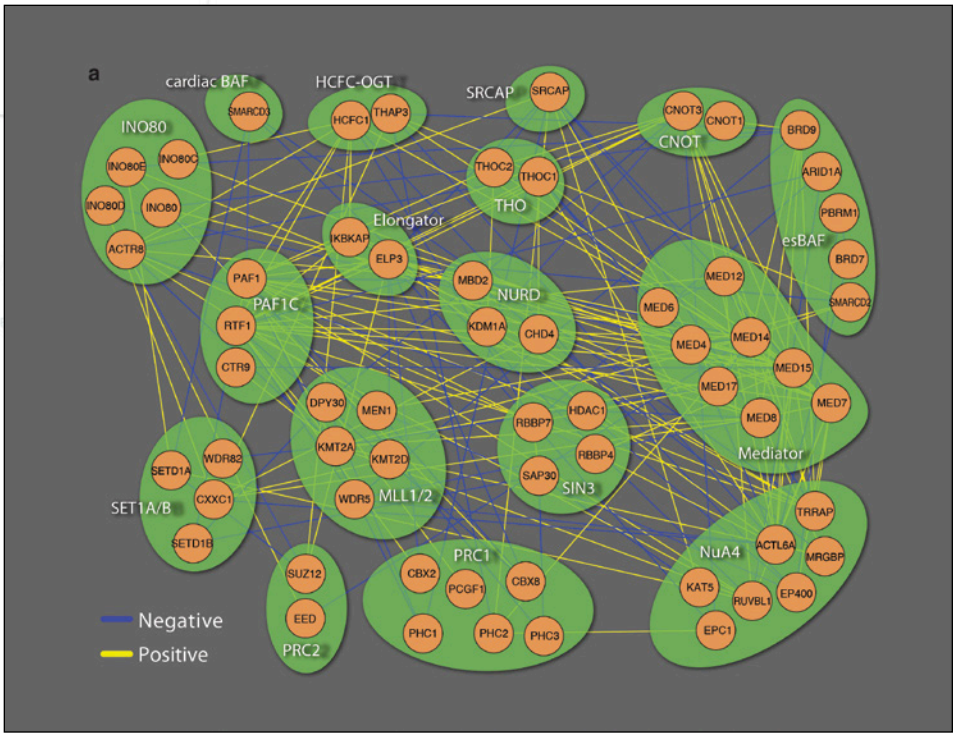
In a 2017 paper in *Nature Methods*, Ideker's team introduced "combinatorial CRISPR-Cas9" for genetic interaction mapping: They used the gene-editing tool CRISPR-Cas9 to knock out single and pairs of genes in high throughput. As a proof of principle, they systematically mutated 73 cancer genes—tumor suppressor genes and cancer-relevant drug targets—one at a time and in all pair-wise combinations in three human cancer cell lines: cervical, lung, and kidney. Two genes interact if their double mutant grows faster or slower than their single mutants would predict.

Ideker's team identified numerous interactions, including 152 synthetic-lethal combinations. Most of these were novel, though some were already known; for example, they rediscovered the BRAC-PARP lethality targeted by olaparib. When they tested eight novel synthetic-lethal combinations by simultaneously drugging both genes, six were experimentally validated.

"Ultimately you'd love to be able to take all 30,000 genes and look at how twiddling all pairs of 30,000 genes affects the function of cells. But that's still too big of an experiment to do, at least with our current state of technology," Ideker says. But, he says, "What's nice about these CRISPR studies is that the speed and coverage of that interaction map is directly coupled to the speed and cost of DNA sequencing. So, if that continues to fall, then it's going to pull the interaction mapping along with it."

Besides experimental advances, computational advances are also helping scientists make headway in genetic interaction mapping. For example, a new algorithm called TransposeNet, introduced in a 2017 paper in *Cell Systems*, leverages network information from model organisms such as yeast to help build human networks. TransposeNet was developed by **Vikram Khurana**, **MD**, **PhD**, assistant professor of neurology at Harvard Medical School and principal faculty at the Harvard Stem Cell Institute, working with a team of computational biologists led by **Bonnie Berger**, **PhD**, Simons Professor of Applied Mathematics and Computer Science at MIT, and including **Jian Peng**, **PhD** (a postdoc at the time).

Khurana studies neurodegenerative disorders in yeast cells and neurons derived from patient stem cells. Yeast may not have brains, but they exhibit critical eukaryotic biology found in specialized cells like neurons, especially when it comes to ancient problems like protein misfolding. In Parkinson's disease, for example, the protein α-synuclein forms clumps (called Lewy bodies) in dopamine neurons. If yeast cells are forced to express α-synuclein—which is not native to yeast—the protein forms toxic clumps similar to those found in Parkinson's disease. Berger and Khurana's team screened the entire yeast genome to identify genes that interact with α-synuclein, either ratcheting up or ratcheting down its toxicity. They came up with 332 genes,



Cancer Network. *By creating double-knockout mutants, researchers were able to identify genetic interaction networks involved in cancer. The green circles represent protein complexes; the orange circles are genes in those complexes. Blue lines indicate negative genetic interactions and yellow lines indicate positive genetic interactions. Reprinted with permission from: Shen JP, Zhao D, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, Licon K, Klepper K, Pekin D, Beckett AN, Sanchez KS. Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions.* Nature Methods *14:573-6 (2017).*

which they then assembled into a biological network. Because the yeast genome has been so well studied, they were able to connect genes based on multiple types of relationships—including genetic interactions and physical ones (for example, protein-protein interactions).

Berger and Khurana then created a "humanized" version of this yeast network using TransposeNet. If you simply convert the 332 yeast genes to human genes using homology mapping, "you fall flat on your face," Khurana says, because there is a dearth of available information about how the human genes interact. TransposeNet solves this issue by using the wiring diagram from yeast to help fill in the wiring diagram for humans. "What we said is if those interactions are conserved and we don't have any of those interactions in humans yet, why don't we use our yeast-to-human algorithm to not just convert a list of yeast genes into a list of human genes, but actually take the entire genetic network in yeast and convert that into the human proteomic space?" Khurana explains.

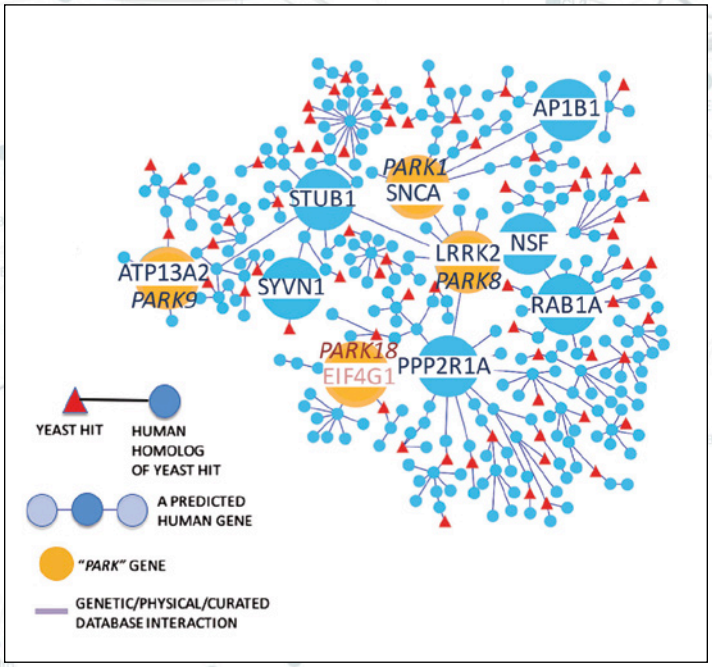
TransposeNet relies heavily on the SteinerNet algorithm developed by **Ernest Fraenkel, PhD**, professor of biological engineering at MIT. This algorithm optimizes network building by prioritizing the most relevant interactions, including pulling in new genes if needed. "We used his algorithm to not just make a network between genes that were already in our list but actually to be able to add genes in, especially if it solves the network in the most efficient way possible," Khurana says.

Remarkably, TransposeNet pulled into the network many known Parkinson's risk genes that don't have homologs in yeast, including the gene for α-synuclein itself. "It was really cool that the algorithm was able to reintroduce the protein of interest and other human proteins that are critically important for Parkinson's disease; we never told the network to do that," Khurana says.

Not surprisingly, the human network was enriched for genes involved in protein trafficking, which is known to go awry in Parkinson's. Unexpectedly, the network was also enriched for RNA binding proteins, which have previously been implicated in ALS but never in Parkinson's disease. RNA binding proteins orchestrate protein translation. "We're very excited about uncovering this new axis of biology," Khurana says. Indeed, when Khurana's team grew neurons in a dish from Parkinson's disease patients with α-synuclein mutations, they found that the neurons had abnormally low protein translation. Moreover, they could reverse this defect by increasing the expression of two genes found to suppress α-synuclein in the original yeast screen.

"Vik Khurana validated these findings in human stem cells, which is unbelievable," Berger says. "It was a real joint effort between the computational biologists, the biologists, and physicians, which I think is really nice. This is how we're going to get the best translation to therapeutics in my opinion."

The network is also being used to discover new Parkinson's disease risk genes. "We believe that there are absolutely going to be additional genes in our network that are contributing to or causing disease," Khurana says. They are now sequencing exomes from patients with Parkinson's disease to look for mutations in network genes.



Parkinson's Disease Network. *Researchers identified genetic modifiers of the Parkinson's Disease gene α-synuclein in yeast, and then used the program TransposeNet to generate a "humanized" network. Genes related to neurodegenerative diseases are enlarged. Red triangles are the original yeast hits; dark blue circles are human homologs of yeast hits; and light blue circles are predicted human genes. Genes are connected if they have a known genetic or physical interaction. Reprinted with permission from: Khurana V, Peng J, Chung CY, Auluck PK, Fanning S, Tardiff DF, Bartels T, Koeva M, Eichhorn SW, Benyamini H, Lou Y. Genome-scale networks link neurodegenerative disease genes to α-synuclein through specific molecular pathways.* Cell Systems. *2017;4:157-70.*

Protein-Protein Interaction Networks: Proteins Teaming Up in Parkinson's Disease

Two proteins that physically interact likely participate in the same biological pathway, so mapping protein-protein interactions can give direct insights into disease. Technological advancements are increasing the pace of mapping; and scientists are approaching genome-wide coverage in at least one human cell at one time point.

One of the largest efforts to map the human protein interactome is BioPlex, led by **Wade Harper, PhD**, and **Steve Gygi, PhD**, professors of cell biology at Harvard Medical School. Harper and Gygi have developed a high-throughput pipeline that uses affinity purification-mass spectrometry (AP-MS). In AP-MS, scientists insert an affinity tag into their protein of interest. This tag then binds to a matching bead, with which the protein can be fished out of a cell along with all its binding partners—which are then identified by mass spectrometry. Although AP-MS can enrich for the protein of interest and its interacting partners, there are also background proteins that associate non-specifically with affinity beads and often dominate the proteins identified by mass

spectrometry. So, the Harper and Gygi groups developed software, CompPASS-Plus, that uses a naive Bayes classifier to help distinguish high-confidence interacting partners from background interactions.

Working at a rate of about 500 proteins per month, this group has now mapped about 10,000 proteins and 120,000 interactions from one human cell line. They have released data for about 7,500 proteins and 50,000 interactions in BioPlex 2.0 (http://bioplex.hms.harvard.edu), which was described in a 2017 paper in *Nature*. People are already using the data, Harper says. "There are several examples where people have taken the network, identified novel interactions, and then made a biological discovery."

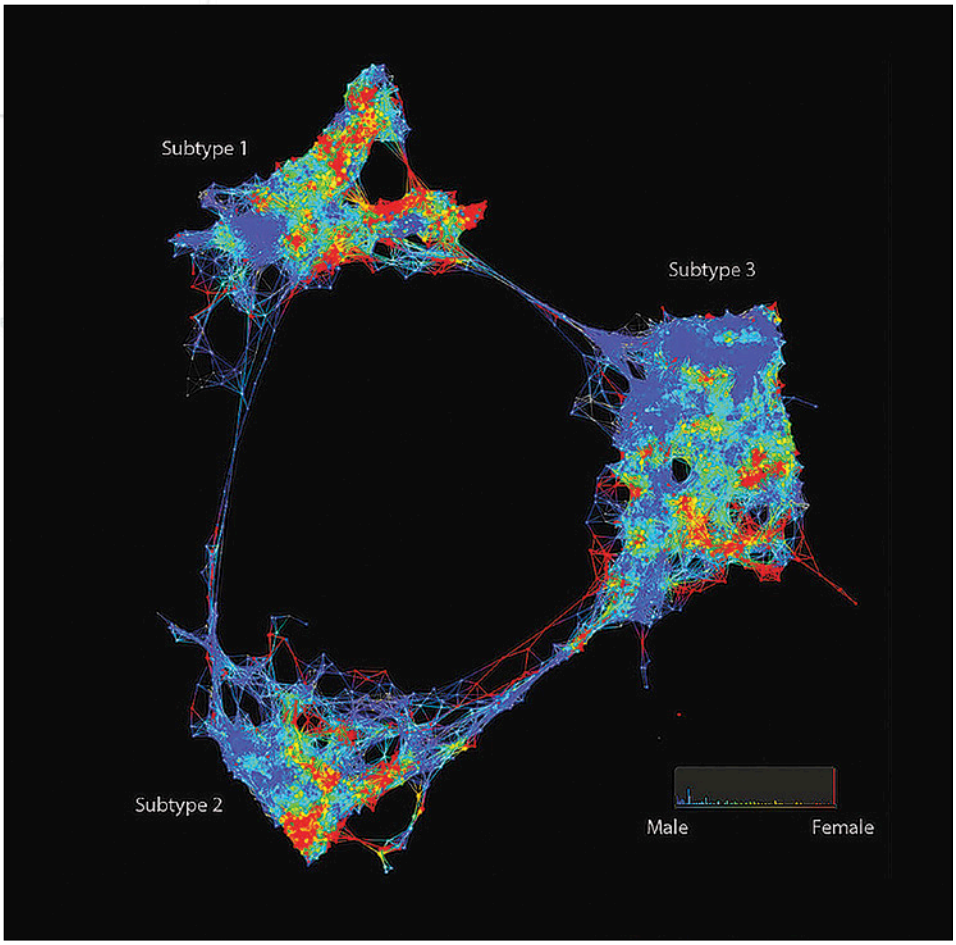
Membrane proteins represent a challenge for high-throughput approaches, as individual complexes require different detergents for proper extraction. Thus, for membrane proteins, the Harper/Gygi team is turning to a new technology called APEX, developed by **Alice Ting, PhD**, professor of genetics, biology, and chemistry at Stanford University. Scientists insert an APEX tag into their protein of interest; when activated, APEX "spray paints" everything in the immediate vicinity (within a few nanometers) with biotin, a small molecule that is widely used to couple proteins, Ting says. These biotin-labeled proteins can be pulled out of the cell with high specificity, reducing false positives. Since the protein complex doesn't have to remain intact, this also reduces false negatives.

Another advantage of APEX is that it can be used to resolve protein interaction networks in space and time. "People tend to view protein complexes as static entities, but, in reality, they're not," Harper says. APEX is incredibly quick—labeling takes as little as 20 seconds—making it possible to map protein-protein interactions dynamically. "With APEX, you can monitor the changes in interaction partners over the time course of a biological process, and get a dynamic picture of what's going on," Harper says. For example, in a 2017 paper in *Cell*, Gygi and colleagues used APEX to map the rapidly changing protein interaction networks of G-protein-coupled receptors following ligand binding.

APEX was central to a second paper that Khurana and colleagues published in *Cell Systems* in 2017 in which they revealed α-synuclein's protein interaction network (the paper was published as a companion paper to the genetic interaction paper). Using APEX, they detected 225 proteins that reside in tight proximity to α-synuclein in rat neurons. Remarkably, the resulting protein interaction network converged on the same Parkinson's risk genes and cellular processes as the genetic interaction network created by TransposeNet. Both highlighted protein trafficking and RNA binding proteins. "One of the really cool things is that the core proteins that had originated in yeast in that other paper were actually interacting with α-synuclein," Khurana says. "So there was this deep relationship between where these proteins are in a cell and the mechanisms through which they exert their toxic effects."

Patient Similarity Networks: Connecting Shared Phenotypes in Diabetes and Cancer

Biological networks aren't limited to the molecular level. Researchers are also building networks at the patient and disease levels. In a patient similarity network, two patients are linked if they share phenotypic similarities. Patient similarity networks can reveal subtypes of disease with similar biological underpinnings,



Diabetes Network. ***A patient-patient similarity network of 2551 diabetes patients reveals three subtypes of disease. Researchers used data from the electronic medical records to build the network. Patients (nodes) are connected if they exhibit similarity across many clinical dimensions (for example, laboratory tests). Patients who exhibited very high degrees of similarity were grouped into single nodes. Colors correspond to gender—with red shades indicating more females and blue shades indicating more males. From: Li L, Cheng WY, Glicksberg BS, Gottesman O, Tamler R, Chen R, Bottinger EP, Dudley JT. Identification of type 2 diabetes subgroups through topological analysis of patient similarity.*** Science Translational Medicine. ***2015;7: 311ra174-311ra174. Reprinted with permission from AAAS.***

which may lead to tailored treatments.

In a 2015 paper in *Science Translational Medicine*, researchers from the Icahn School of Medicine at Mount Sinai used patient similarity networks to identify three subtypes of type II diabetes. **Joel Dudley, PhD**, associate professor of genetics and genomic sciences at Icahn, and his colleagues culled 73 objective clinical measures—such as height, weight, and blood panels—from the electronic medical records of 2,551 diabetes patients. They used a dimensionality reduction technique to compress these 73 variables into a more succinct representation; and then clustered patients based on similarity.

The resulting network had three distinct clusters of patients, which the researchers then attempted to characterize. They compared the clusters' clinical characteristics, as well as their comorbidities and genotypes—factors held out of network building. Patients in subtype one were characterized by classic type II diabetes features—including obesity, high blood sugar, and kidney and eye disease. Patients in subtype two were thinner, had more cancer and cardiovascular disease, and had polymorphisms in immune genes. "I call them skinny immune diabetics," Dudley says. "They appear to have an immune- or inflammatory-driven diabetes that differs from the classical metabolic dysfunction." Patients in subtype three had high levels of mental illness as well as cardiovascular disease. Psychiatric medications are known to increase the risk of diabetes, and could partly explain this subtype.

When Dudley presented these data to physicians, the subtypes resonated with them. "Hindsight's 20-20, but when we showed them this they were like 'Oh, I know that type of patient,'" Dudley says. "They said it over and over again."

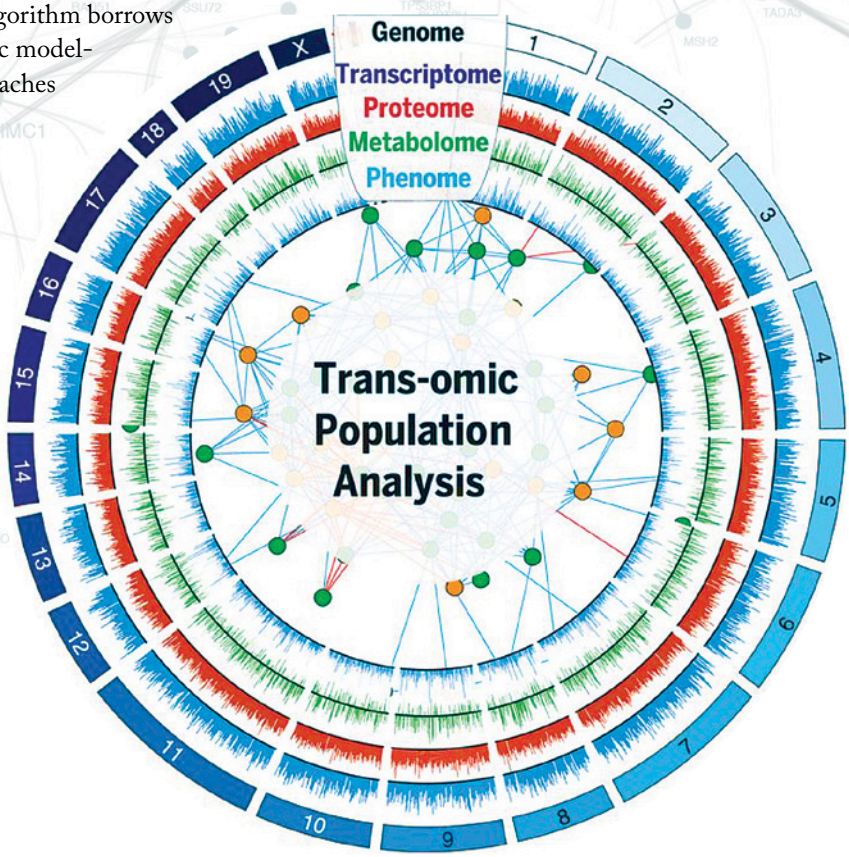
One limitation is that the data capture just one snapshot in time. "I think a lot of the networks we've built today are just glimpses into the real networks in our bodies," Dudley says. To get a fuller picture, Dudley's team is working on building networks using longitudinal patient data. "What's becoming clear is that networks are super dynamic, super ephemeral. And we need lots of different perturbations to really understand the actual logic and topology of the networks," he says.

When building patient similarity networks, most scientists cluster patients first and then look at the underlying characteristics of the clusters. But **Teresa Przytycka, PhD**, a senior investigator in the Computational Biology Branch of the National Center for Biotechnology Information at the NIH, has taken a different tactic. Her team uses a probabilistic algorithm that builds the patient network using phenotypic similarity data (e.g., gene expression or survival data) and genetic features (e.g., single-nucleotide polymorphisms or copy-number variations) simultaneously. This way, patients are not forced into a single subtype but are allowed to reflect a mixture of subtypes.

The algorithm borrows from topic modeling approaches in text mining. Topic modeling attempts to classify documents into overarching topics—say sports, politics, culture, and science—by analyzing the set of words in each document. Similarly, Przytycka's team attempts to classify patients into subtypes by analyzing the set of genetic perturbations in each patient. Just as a document may be classified as both politics and science, a patient may also be classified as a mix of subtypes. "This is an example of taking a method from one field and applying it to another field. We can take software that experts have been working on for years and then repurpose it for molecular biology questions," Przytycka says. When applied to glioblastoma, her team found that the disease has only three, not four, subtypes; what had previously been deemed a fourth subtype was in fact just a mixture of two other subtypes.

Integrated Networks: The Big Prize

In isolation, each network type gives just one piece of the puzzle. The "big prize" in network biology will be to integrate them, says **Ruedi Aebersold, PhD**, professor



Trans-omic Network. *Researchers created an integrated network using five layers of network data—genetic, expression, protein, metabolic, and phenotypic—pertaining to fat metabolism in mice. The combined analysis of all layers together provides additional information not yielded by any single omics approach. From Williams EG, Wu Y, Jha P, Dubuis S, Blattmann P, Argmann CA, Houten SM, Amariuta T, Wolski W, Zamboni N, Aebersold R. Systems proteomics of liver mitochondria function.* Science. *2016;352(6291):aad0189. Reprinted with permission from AAAS.*

of systems biology in the department of biology at Swiss Federal Institute of Technology in Zurich (ETH Zurich). "They're all views on the same cell from a different angle, so we'd like to be able to integrate various types of networks into a more global model." It's an enormous challenge; and solving it will require computational biologists and experimental biologists to work together, he says.

In a paper in *Science* in 2016, Aebersold's team (in collaboration with the group of **Johan Auwerx, MD, PhD**, from Ecole polytechnique fédérale de Lausanne [EPFL]) presented an analytic pipeline to measure and combine five layers of network data—genetic, expression, metabolic, protein, and phenotypic. They measured about 25,000 transcripts, 2600 proteins, and 1000 metabolites from 40 different strains of mice with well-characterized genetics. The project team fed the mice a high-fat or low-fat diet; and then measured phenotypic responses such as weight change, glucose tolerance, and the presence of fatty liver disease. The researchers then correlated elements across different network types. For example, they measured the correlation between a gene's RNA expression level and the levels of the protein translated from that RNA—and surprisingly, these weren't always tightly linked. They assembled all the data into a single network where the nodes were genes, transcripts, proteins, metabolites, or phenotypes, and the edges were the correlations between these. The resulting network gave novel insights into the role several proteins play in metabolizing fat.

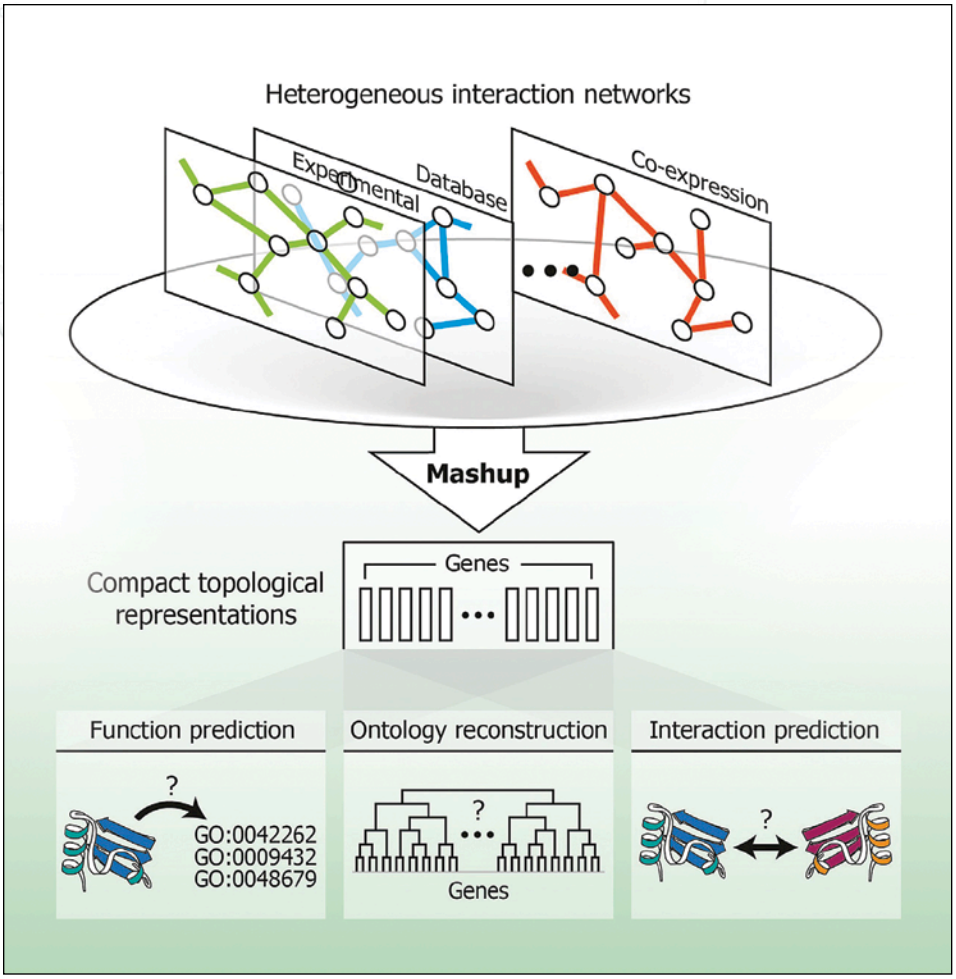
With Peng, Berger has also built a computational pipeline for combining networks called MashUp (http://mashup.csail.mit.edu). The connectivity pattern of each node in a network is incredibly complicated—but this information can be compressed into a simpler representation in the same way that Google's PageRank condenses a website's connectivity patterns into a simple ranking. MashUp extracts this information from multiple networks for each gene and then integrates it into one measure of global connectivity that informs how the gene relates to other genes in the networks. "We generate compact representations for the topology of each node in its network and then integrate that using off-the-shelf machine learning methods," Berger explains.

In a 2016 paper in *Cell Systems*, Berger's team showed that information extracted and combined with MashUp can be used for tasks such as automated gene function annotation with substantial improvements over state-of-the-art methods. "I think this paper goes a long way to solving the issue of how to integrate multiple network topologies," Berger says.

Toward Precision Medicine

If you can connect all the networks into one integrated picture of a complex disease, that's a first step toward using them to provide medical care that's tailored to individual patients. "We're all working toward precision medicine," Ideker says.

He envisions a "clinic of the future" where a computer simulates diseased cell or tissue, informed by all the available interaction data. "You would load a patient's particular mutations and environmental conditions onto that model and you would compute the drug that best returns that model to its normal state," Ideker explains. His lab is already working on such a model for a cancer cell. "Of course, we're not going to get there tomorrow. But it's important to have the vision," Ideker says. "We're going after this vision already. Time will tell how fast we can push it." □



Melded Network. *In Mashup, the high-dimensional topological patterns in individual networks are canonically represented using low-dimensional vectors, one per gene or protein. These vectors can then be plugged into off-the-shelf machine learning methods to derive functional insights about genes or proteins including protein function prediction, gene ontology reconstruction, and genetic interaction prediction. Reprinted with permission from: Cho H, Berger B, Peng J. Compact integration of multi-network topology for functional analysis of genes.* Cell Systems. *2016;3:540-8.*

DEEP LEARNING AND THE FUTURE OF BIOMEDICAL IMAGE ANALYSIS

BY ALEXANDER GELFAND

Revolutionary technological advances in the areas of autonomous vehicles, speech recognition, cybersecurity, and earthquake prediction all depend on the family of artificial intelligence techniques known as deep learning.

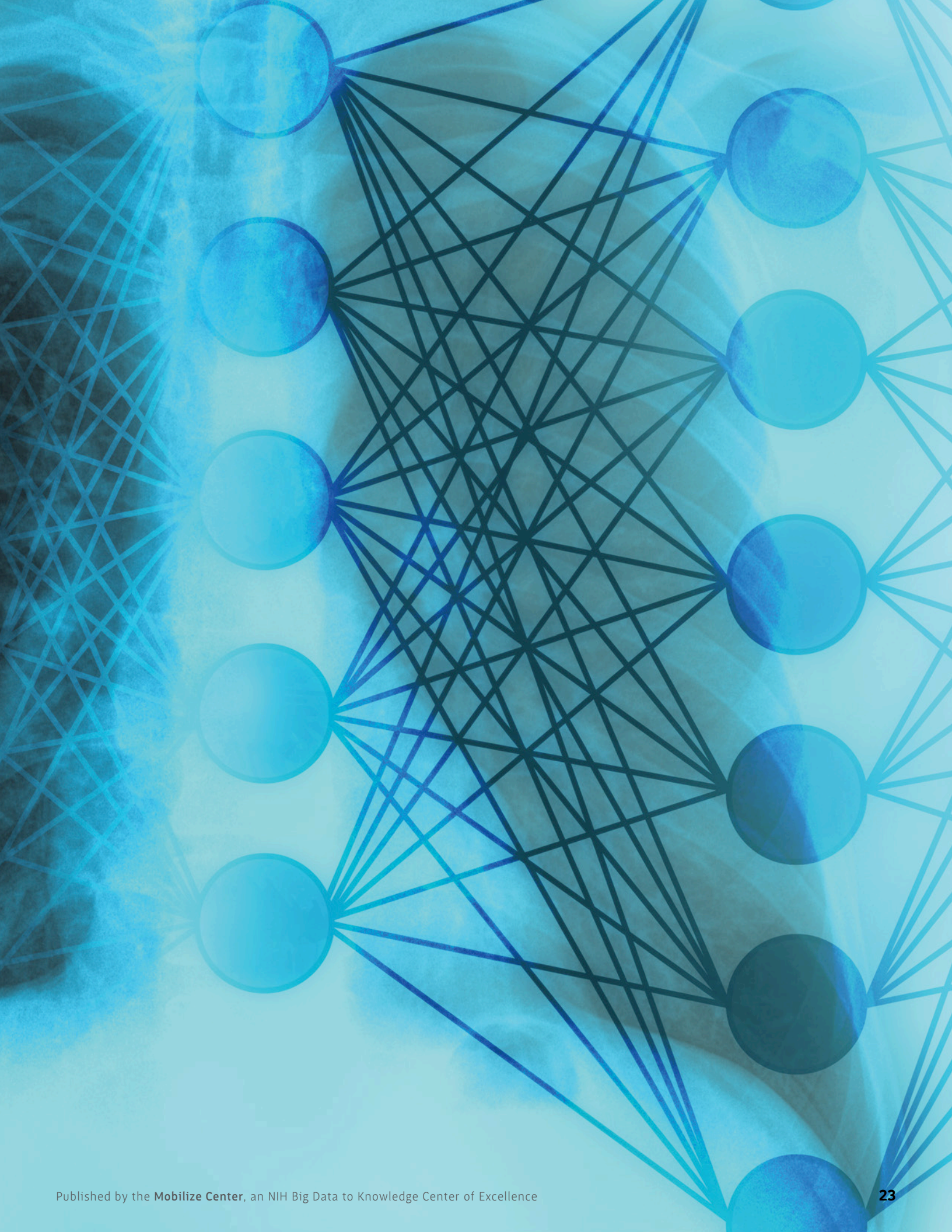
But no field stands to benefit more from this approach than biomedical image analysis, a painstaking task that currently falls to highly trained radiologists and pathologists. Imagine, for example, a computer program that can detect a suspicious mass in a mammogram or identify a handful of abnormal cells in a biopsy slide.

The application of deep learning algorithms to biomedical image analysis is still in its infancy. Yet researchers around the world are already achieving uncanny results, and it is only a matter of time before their as-yet-experimental models enter the clinic. Once that happens, proponents say, deep learning will enable earlier and more accurate disease detection, allow more precisely tailored treatment plans, and ultimately improve patient outcomes.

"In the next two to five years, I see for the first time the possibility that the engineering tools that we're developing will actually affect the clinic and start to advance medicine," says **Hayit Greenspan, PhD**, head of the Medical Image Processing and Analysis Lab at Tel Aviv University and co-editor of the book *Deep Learning for Medical Image Analysis*.

Yet before deep learning can realize its potential to transfigure biomedicine, it has some hurdles to leap. For one thing, the technology is still fairly limited and is currently better suited to performing rote tasks rather than more advanced diagnostic ones. For another, there is some question regarding just how much artificial intelligence patients and doctors will accept in the clinic. And finally, there's an urgent need for more training data to teach the deep learning models how to do their jobs—a problem that researchers are currently addressing in multiple ways.

Ultimately, how these issues are resolved will determine precisely



how deep learning will enter the realm of clinical medicine. Nonetheless, if recent experimental studies are any indication, it's ultimate impact is virtually assured—a question of how and when, rather than if.

Computer, Teach Thyself

Deep learning relies on sophisticated statistical models known as neural networks. Inspired by the human brain, these consist of virtual neurons—what **Thijs Kooi**, a PhD candidate in the Diagnostic Image Analysis Group at Radboud University Medical Center in the Netherlands, describes as "rudimentary elements of computation." In a deep neural network, these are organized in multiple data-processing layers. Each layer transforms a piece of input and passes it to the next layer in such a way that the model can eventually learn to master complex tasks that it was never programmed to handle.

Traditional machine learning algorithms must be told which image features to consider when classifying a tumor as malignant or benign. But if you feed enough images of malignant and benign tumors to a neural network—or more precisely, to a deep convolutional neural network (CNN), a species of deep learning model that is particularly well suited to image analysis—it will eventually learn to distinguish between them on its own.

Robot Doctors? Not So Fast: Short-Circuiting the Hype Cycle

In a paper published in the journal *Nature* in February of 2017, **Andre Esteva** and **Brett Kuprel**, PhD candidates in the department of electrical engineering at Stanford University, reported that a CNN they trained on more than 1.4 million images was ultimately able to detect and classify various forms of skin cancer as accurately as 21 board-certified dermatologists.

Similarly, researchers at Google reported in 2016 in the *Journal of the American Medical Association* that they trained a CNN to diagnose diabetic retinopathy—an eye disease that afflicts almost one third of all diabetes patients and constitutes a leading cause of blindness—as accurately as seven board-certified ophthalmologists. More recently, several of the same Google researchers trained a CNN to match and even exceed the performance of a pathologist when identifying slide images of breast cancers that had metastasized to a patient's lymph nodes.

Other groups, meanwhile, have used deep learning to identify signs of Alzheimer's in MRI scans of the brain, to detect lung cancer, and to spot musculoskeletal abnormalities in bones and joints.

With successes like these, it might seem as if deep learning is on the cusp of rendering obsolete the highly educated human beings who are currently responsible for analyzing medical images. And advocates do anticipate that their autodidactic algorithms will soon undertake at least some of the tasks currently performed by flesh-and-blood doctors.

"I think it should be possible to replace routine image reading tasks in the next 5 to 10 years or so," says Kooi.

But today's physicians need not worry about their job security just yet. For one thing, claims of deep learning models matching—or besting—the performance of human beings can be misleading, Greenspan says.

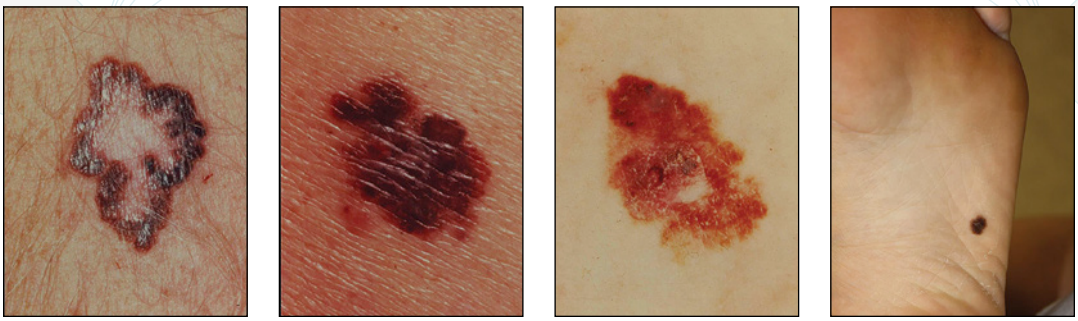
Often, such horse races are conducted by inviting radiologists or other expert human readers into the lab and showing them the same simplified 2-D images that are fed to a model. But this does not resemble the real-life workflow of a radiologist; and under those circumstances, a human reader may prove far less reliable than he or she might in the clinic, making the algorithmic competition look far better by comparison.

It's possible that deep learning is approaching the peak of a "hype cycle," says **Daniel Rubin, MD, MS**, associate professor of biomedical data science, radiology, and medicine, who runs the Stanford Quantitative Imaging Laboratory. He collaborates on a variety of projects that employ a broad array of machine-learning methods and worries that deep learning is drawing attention away from other technologies that are still useful.

Deep learning algorithms have proven remarkably adept at standard radiological and pathological tasks such as segmentation, detection, and categorization. Yet as Greenspan explains, all of these are essentially problems of classification. For example, a deep neural network can be assigned the task of analyzing an x-ray, CT scan, or MRI at scales ranging from single pixel to region of interest or entire image, estimating the probability that it belongs to a particular class and labeling it accordingly: organ or surrounding tissue, normal or abnormal, cancer type A or cancer type B.

But as Rubin points out, much of what doctors do goes well beyond pattern recognition and image classification. It involves a complex combination of knowledge, reasoning, and inference. And while deep learning might eventually capture all of that, right now, it falls far short.

It is therefore likely that deep learning models, which do not grow bored or fatigued when forced to examine scads of mammograms or slide



Recent work to diagnose and classify skin cancers using deep learning has proven remarkably successful. Image courtesy of the website of the National Cancer Institute (https://www.cancer.gov).

images, will initially be applied to relatively mundane tasks. Greenspan points out that this will allow physicians to become accustomed to the technology while enhancing their productivity and accuracy, freeing them to deal with more subtle problems, even as deep learning researchers gradually hone more advanced applications.

Shadi Albarqouni, PhD, a postdoctoral research associate at the Technical University of Munich, and colleagues recently trained a CNN to decompose chest x-rays in such a way that bony structures like the ribs and spine, which can obscure the soft tissue of the lungs, are eliminated from the picture. This would allow radiologists to more easily focus on areas of interest and perceive soft tissue abnormalities, thereby improving their chances of making a correct diagnosis.

And in a paper published in 2016, Albarqouni and others trained another CNN to analyze fluoroscopic x-ray images and identify the catheter electrodes that surgeons insert into patients during electrophysiology procedures, tagging them with colored labels and estimating their depth to enable precise placement.

Beyond such supporting roles, however, it is unclear exactly to what extent physicians—and their patients—will accept AI in the realm of healthcare.

Rubin himself has used CNNs to identify and grade the brain tumors known as gliomas in digital images of histopathology slides, and to identify and localize masses in mammograms.

Yet as he points out, "patients want a human being in the loop in their care." So does the law, which requires that human beings, not algorithms, assume liability for medical decisions. "Can it be legally acceptable to have a computer practice medicine and replace the decision-making of a person?" he asks.

The Black Box Problem

Physicians themselves may also be uncomfortable with the results of deep learning because CNNs are black boxes. While they can determine which features are most useful for discriminating between different classes of images (e.g., tumors versus benign masses), the models do not reveal which of those features they rely upon, or how, precisely, they arrive at their decisions—for example, if a tumor is malignant or not.

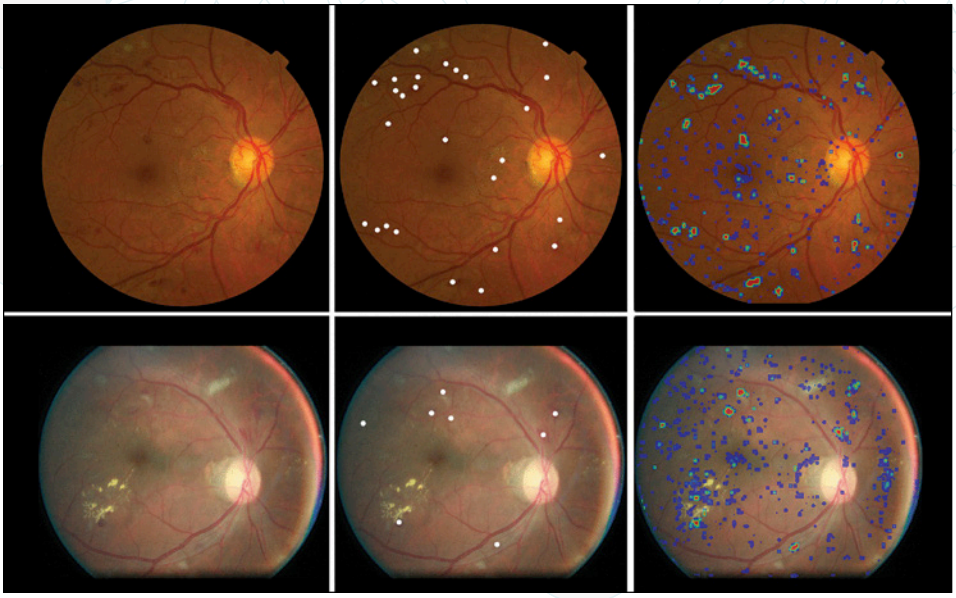
"The nature of these models is such that we give them a raw framework for how we think a problem works, and they fill in all the details," says Kooi, who develops deep learning models capable of detecting breast cancer in mammograms. "How it fills in all the details is something we don't really have a lot of control over."

Researchers are working on ways of peeking inside the models to understand how they select discriminant features. For example, a group of Stanford graduate students led by **Avanti Shrikumar**, a PhD candidate in computer science, recently developed an algorithm called DeepLIFT that attempts to determine which features are important by analyzing the activity of a model's neurons when they are exposed to data. A team of engineers at the Israel-Technion Institute of Technology have devised a method of visualizing the neural activity of a network that resembles what one sees in fMRI of the human brain. And Rubin recently published a paper in which he and his colleagues trained a CNN to distinguish between benign and malignant breast tumors, and then used a visualization algorithm, called Directed Dream, to heighten and exaggerate specific details in order to maximize the images' scores as either benign or malignant. The resulting "CNN-based hallucinations" effectively show how the CNN learns clinically relevant features, lending credibility to the results.

But much work remains to be done on this front. And Rubin suspects that while doctors might be willing to accept deep-learning input for straightforward screening tasks—so long as the network's track record is strong—they would be far less likely to accept something more complex and consequential like a diagnosis from a model whose inner workings remained a mystery.

"Physicians will not accept the output of a decision support system that does not also provide explanations for its answers," he says.

For all these reasons, even if deep learning were to reach the point where it rivaled



To diagnose diabetic retinopathy, doctors examine photographs of the back of the eye, or fundus, for hemorrhages. Deep learning models can be trained to recognize such signs as well. The leftmost column shows unannotated fundus images from a large dataset hosted by Kaggle.com, an online platform for predictive modeling and analysis competitions, that researchers at Radboud University in the Netherlands used to train and test a CNN. The middle column shows the same images with hemorrhages marked by expert human annotators. The rightmost column shows the output of the CNN. © 2016 IEEE. Reprinted, with permission, from van Grinsven MJJP, van Ginneken B, Hoyng CB, Theelen T, S´anchez CI., 2016. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. IEEE Trans Med Imaging *35 (5), 1273-1284 (2016).*

human intelligence, it wouldn't necessarily represent the end of human doctors.

The Data Labeling Dilemma

Yet another obstacle must be surmounted if deep learning is to achieve its full potential: the quantity of the data that this particular flavor of AI requires in order to work its magic.

Deep neural networks can do more automatically than any prior class of machine-learning model. Feed them enough properly labeled data, and they will learn to perform a given task without human intervention.

But the phrase "enough properly labeled data" is a crucial one. A computer's astonishing capacity to learn without human intervention comes at a price; namely, the need for a massive amount of annotated training data. A CNN, for example, may be capable of learning to distinguish between benign and malignant tumors all by itself. But to do so, it must first be fed thousands or perhaps millions of images that have already been correctly labeled benign or malignant, a process known as supervised learning.

"This is the 'no free lunch principle,'" says **Alex Ratner**, a PhD candidate in computer science at Stanford. Deep learning, he explains, may do much more on its own than other machine learning methods; but "it needs more training data to make up for that extra complexity in the models."

Unfortunately, large-scale annotated databases of biomedical images can be hard to come by, and having expert human annotators create new ones from scratch is laborious, costly, and time-consuming. Access to sufficient labeled training data is, therefore, a significant obstacle.

"We work closely with hospitals and radiologists to give us annotated data," says Greenspan, who has used CNNs to detect metastatic liver cancer in 3-D CT scans, segment multiple sclerosis lesions in MRI images, and label pathologies in x-ray images, among other things. "Collecting the necessary data for these tasks is a slow and demanding process."

In addition, says Kooi, "These models are still relatively stupid." In particular, they don't do nuance very well: Having learned to spot obvious examples of common cancers by sifting through a particular training dataset, for instance, they may stumble when confronted with rare or unusual ones, or with anomalies they have never seen before.

Dealing with Nuance: Data Augmentation

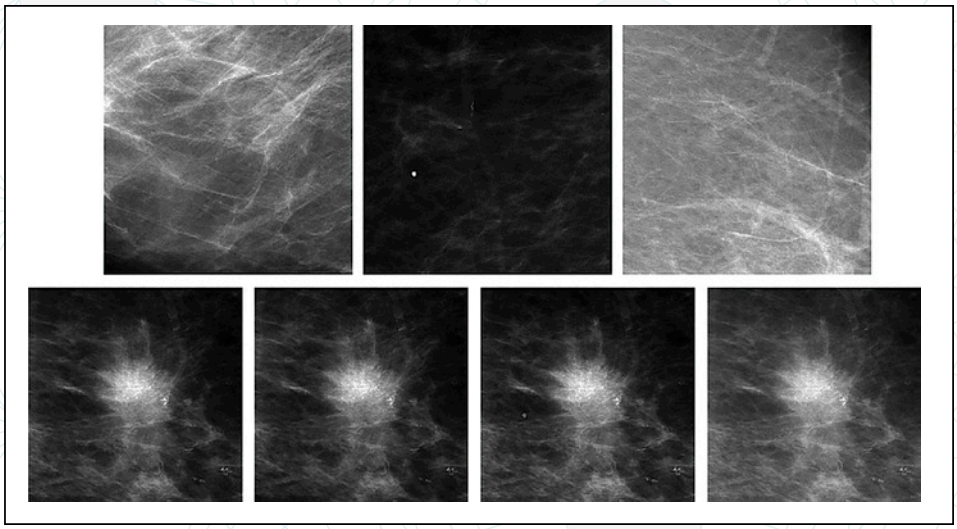
To some extent, researchers can compensate for a lack of labeled training data by using a technique known as data augmentation. This involves transforming some of the training data upon which the deep learning model hones its skills (e.g., rotating images, altering their color, simulating jitter, etc.), thereby preparing it for the kinds of variations and artifacts it might encounter when it is asked to process previously unseen data.

In a paper that appeared in *Medical Physics* this past March, Kooi and his colleagues successfully used a CNN to distinguish benign cysts from malignant masses in standard digital mammograms, achieving results comparable to those attained with a cutting-edge imaging technique known as spectral mammography. They did so in part by processing some of their training data to mimic the natural variations in the amount of tissue that can surround a breast tumor, thereby altering its appearance in a mammogram.

Kooi's CNN was required to analyze only small patches extracted from much larger images, however—patches that another model, known as a candidate detector (which was not itself a deep neural network), had already singled out as containing regions of interest.

Radiologists, on the other hand, typically examine complete mammograms, viewing suspicious areas in the context of the entire image. They also track changes in a patient's scans over time, and note potentially significant symmetries and asymmetries between the left and right breasts. And they consider a whole wealth of information—a patient's lab results, her medical history, her age and demographics—that is not contained within the images themselves.

Putting all of this into the hopper leads to better informed and more accurate diagnoses. Kooi and others are trying to incorporate such diverse sources of information into deep neural networks, in part by integrating them with other computational methods; but they are not there yet.



Data augmentation is used to enhance small collections of training data. Here, Thijs Kooi and his colleagues at Radboud University in the Netherlands employ it to generate variations on the images used to train a CNN to distinguish between breast cancer tumors and benign cysts. The three images in the top row are of normal breast tissue. The leftmost patch in the bottom row, on the other hand, contains a mass or cyst. Superimposing the three normal images from above over the abnormal one produces the remaining images in the bottom row, simulating the different amounts of tissue that might surround a lesion in a mammogram. Reprinted with permission from Kooi T, van Ginneken B, Karssemeijer N, den Heeten A, Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network, Medical Physics *44:3 (1017-1027) 2017.*

Relying on General Knowledge: Transfer Learning

Data augmentation represents one way of addressing the problem of limited training data. So-called transfer learning represents another: a CNN can initially be trained on a large dataset of standard images (dogs, cats, planes, umbrellas), and then retrained on a smaller dataset of biomedical ones (brain scans, chest x-rays, pathology slides). The idea is that the network will learn general, broadly applicable image features from the larger dataset, and transfer or apply that acquired knowledge to the smaller dataset, which will fine-tune it for a particular task such as segmenting organs or detecting lesions.

"The intuition behind transfer learning is that a radiologist doesn't develop a whole new visual cortex every time he learns a new task, but relies on stuff that he already knows—and we can do the same," says Kooi. "We can take a network that was trained on discriminating cats, for instance, then adapt the network to medical tasks."

Esteva and Kuprel, for example, pre-trained their CNN on a set of approximately 1.28 million images, comprising 1,000 object categories, culled from ImageNet, a massive online visual database. They then retrained the model on 129,450 dermatologist-labeled clinical images drawn from clinician-curated online repositories, and from the Stanford University Medical Center. The end result: a model that performed as well as a clutch of human doctors.

In his 2017 *Medical Physics* paper, Kooi took the concept of transfer learning a step further: Instead of starting with generic images, he pre-trained his deep neural network on a large dataset of screening mammograms, essentially teaching the model to distinguish tumors from non-tumors—a medical task that was related to, but not quite the same, as the one he was really interested in. He then retrained the network on a much smaller dataset of diagnostic mammograms, and had it learn how to discriminate between tumors and cysts.

"Often, people train their models on ImageNet, and then fine-tune them with a medical data set," Kooi says. "My argument is, 'It's always better to train the model using a task that is more related to the problem that were trying to solve.'"

In the end, his model nearly matched the performance of a system that used a different kind of statistical model—one that relied on features selected by human beings, rather than on deep learning—along with a more advanced form of mammography.

Outsourcing Annotation

There are many other ways of dealing with the paucity of biomedical training data, some of which take a creative approach to annotation itself. In a paper published last year, for example, Albarqouni and colleagues combined the ground-truth of expert annotations with the crowd-truth of nonexpert ones.

The goal was to improve the performance of a CNN that was trained to detect instances of mitosis, or cell division, in breast cancer biopsy slides. These visible signs of mitosis, known as mitotic figures, appear as small black dots under the microscope, and represent an important criterion for determining the aggressiveness of a tumor—and hence for establishing a patient's prognosis and course of treatment.

The model was trained on expertly annotated images from only eight patients. But when it came time for the network to label previously unseen images, Albarqouni and his team introduced a new twist: whenever the model predicted that an image was more than 90 percent likely to contain mitotic figures, it cropped the region of interest and sent the resulting patch for annotation by at least 10 non-experts via a crowdsourcing platform. These nonexperts, who lacked any medical experience, were given a brief training session and a quiz designed to determine their accuracy. They were then asked to label the patches they received: Were the little black dots singled out by the CNN mitotic figures, or not?

Different users often arrive at different judgments, resulting in conflicting or noisy labels. To resolve such differences, Albarqouni and his colleagues built an "aggregation layer" that collected everyone's annotations and arrived at a consensus label for each patch through majority voting, with each user's vote weighted according to their accuracy.

The results of that vote were returned to the deep learning model, which took the crowdsourced labels into account in its next round of predictions—a crowd-driven fine-tuning process that effectively boosted the network's performance, as measured by the ratio of true positives to false positives, by 3 percent.

In a subsequent project, Albarqouni gamified the crowdsourcing component. He and his colleagues transformed the patches into 3-D stars whose shape, size, and color corresponded to the likelihood that they contained mitotic figures. They then asked users to play a game in which they used a virtual plane to collect the best candidates. This "playsourcing" platform performed 10 percent better than its non-gamified counterpart, an improvement that Albarqouni attributes to the motivating influence of gameplay.

"The player is trying to get a better score," he says; and that translates into a performance boost for the model. It's a win-win.

Automating Everything Noisily

One group is taking the crowdsourcing idea one step further: They're generating (and de-noising) cheaper and messier training data with rules that people write and machines apply. It's a method called data programming that was developed by Alex Ratner and others working in the lab of Stanford computer scientist **Christopher Ré, PhD**. Rather than requiring domain experts or non-experts to hand-label large datasets for training purposes, data programming allows them (or their friendly neighborhood coders) to write small snippets of code that encapsulate the heuristics and rules of thumb that they would use to annotate the data themselves. Those bits of code—called labeling functions—are then used to develop a generative model that can automatically label large

quantities of data for training purposes.

Because these labeling functions may overlap and conflict, the labels they produce are inaccurate, or noisy. (The process of training a model using such noisy labels is known as weakly supervised learning, or weak supervision.) But Ratner and his colleagues, who have developed an open-source data programming platform called Snorkel, use a variety of computational methods to compensate. The large volumes of noisily labeled data pumped out by the generative models created with Snorkel, which are not deep neural networks, can therefore be used to train high-performing discriminative models, which are.

Most of their early efforts involved building text-based datasets. But members of the Ré lab have begun applying data programming and other weak supervision techniques to images as well. And some of their most interesting work involves both.

Ratner, for example, has been working with Rubin and Stanford radiologist **Lawrence Hoffman, MD**, on various projects involving radiological images and their accompanying text reports. The images have not been labeled, but the information required to do so—a physician's clinical judgment that a bone tumor is benign or malignant, for example, or that an arterial blockage has been cleared—is buried in the reports.

Ratner is therefore working on ways of writing labeling functions in Snorkel that can "read" radiology reports and extract labels that can be used to train an image model such as a CNN, enabling it to classify radiological images without the benefit of any text whatsoever. This "cross-modal" approach has succeeded with test data, and will soon be deployed on real clinical data.

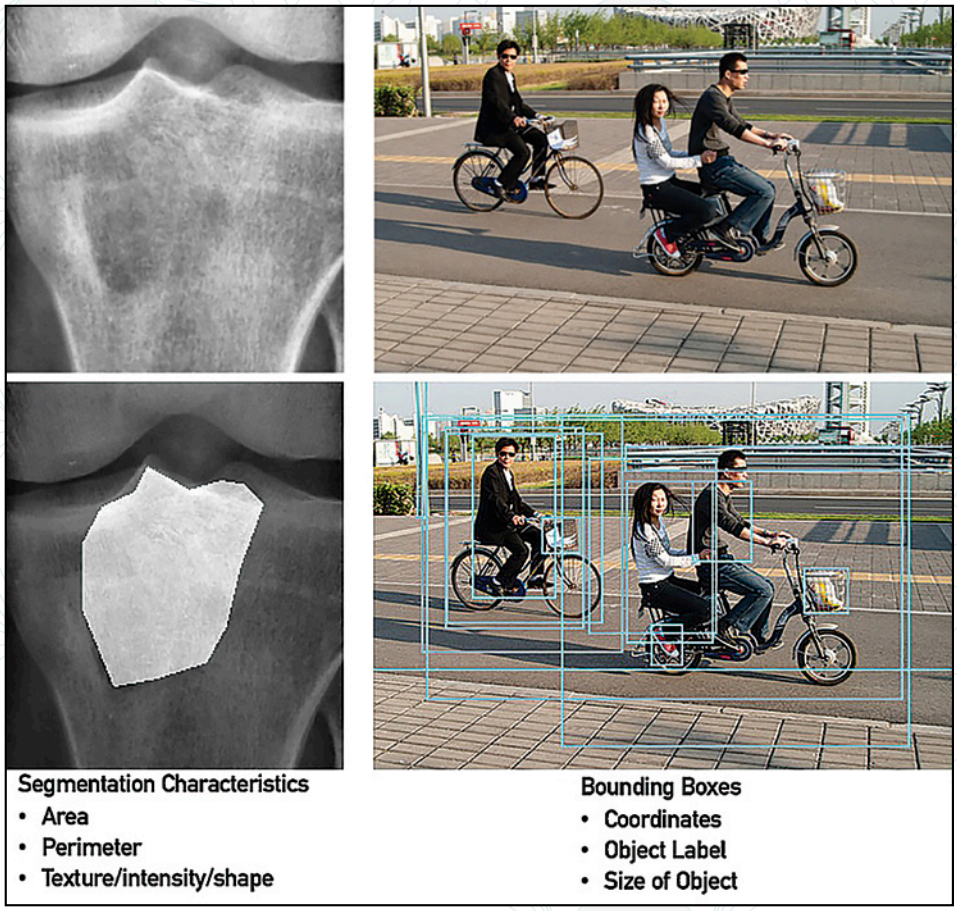
"If this works, then the model you've trained could look at an image before the radiologist has actually dictated the report, and come up with a classification," says Ratner.

Paroma Varma, another doctoral candidate in the Ré lab, and colleagues have developed a different software platform called Coral to apply the idea of weak supervision directly to images and video. In a recent paper, she and one of Rubin's PhD students, **Darvin Yi**, had Coral label tumors in mammograms as either malignant or benign, and used that data to train a CNN to distinguish between unlabeled examples of the two. Remarkably, this Coral-trained CNN proved almost as accurate as a CNN that had been trained on a small hand-labeled dataset.

Varma and another member of the Ré lab, PhD candidate **Braden Hancock**, are now attempting to crowdsource the act of creating labeling functions. In one proof-of-concept experiment, they posted images to the crowdsourcing platform Mechanical Turk and asked users not only to label them, but also to explain their reasons for doing so. They then used a language tool known as a semantic parser to convert those natural language explanations into code. The result: instant labeling functions, automatically generated from standard English. The images Varma and Hancock used were not medical ones, but if Albarqouni's crowdsourcing work is any indication, the approach certainly holds promise.

Other automated solutions to the training data dilemma are also under development. Greenspan, for example, has been using deep learning models known as Generative Adversarial Networks (GANs) to generate synthetic training data using small training sets. Her work is still in the experimental phase, but if successful, it could enable one deep learning model to produce the data required to train another deep learning model.

Given the pace at which the field is moving, the application of deep learning to biomedical images is likely to keep plenty of doctors and computer scientists busy for the foreseeable future. Radiologists will be presented with more and more information to factor into their clinical decisions as artificial intelligence gradually enters the scene. And computer scientists, for their part, will continue to grapple with the challenges presented by data-hungry deep learning models and the noisy medical data required to feed them. ☐



Paroma Varma and her colleagues in Christopher Re's lab have developed a platform called Coral to help users label large training datasets of images and video for deep learning purposes. Such datasets can be used to teach a deep learning model how to recognize images of people on bicycles—or how to segment a potential tumor on a bone x-ray. Courtesy of Paroma Varma.

BY CHAND T. JOHN, PhD

WHAT VALUE COULD FRACTALS ADD TO BIOMEDICAL IMAGE ANALYSIS?

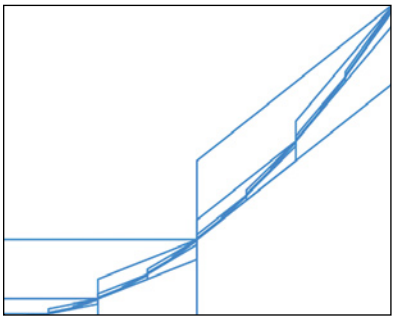
We collect large amounts of biomedical image data, hoping to glean insights into our biological world. While deep learning has become popular for finding features that, for example, distinguish between benign and malignant tumors in biomedical images, how these features relate to conclusions we care about remains a mystery hidden in a labyrinth of neural networks.

Fractals, "self-similar" shapes whose parts resemble the overall shape itself, better connect biomedically relevant properties to their internal parameters. A fractal has properties like fractal dimension, a measure of how its complexity changes across scales, and lacunarity, a measure of sparsity and non-uniformity, that relate to patterns of interest in biomedical images. For example, fractal models of the lung have revealed correlations between tumor presence, higher fractal dimension, and lower lacunarity. Fractal modeling of brain cancers has allowed determination of tumor stages using fractal dimension. These medically relevant properties of a fractal are closely tied to its self-similar structure, which itself emerges from the fractal's defining parameters. In this way, fractals offer a tantalizing strength in connecting biomedically relevant properties back to their internal mathematical parameters in a way that deep-learning models currently do not.

While fractals lend insight into their own internal structure, it's tempting to ask whether fractal models can go even deeper. Could a fractal's parameters somehow relate to other meaningful aspects of an image's geometry, similar to how Newton's laws relate geometric parameters of planetary trajectories to the geometry of underlying gravitational force fields?

One type of fractal offers a surprisingly tangible set of parameters: smooth curves that resemble the fractal's own shape. Smooth fractals are curves such as parabolas, which, as it turns out, are fractals themselves. For example, a pair of mappings called affine maps, carefully chosen, can squeeze and stretch a box repeatedly in a way that causes the resulting shape to converge toward a parabola.

Starting with multiple smooth fractals, combining their affine maps into one aggregate mapping, and repeatedly applying that aggregate mapping to a box, results in a shape that, despite its complexity, is completely characterized by those smooth curves with which we started.



Starting with a square and repeatedly applying two affine transformations leads to a set of parallelograms that converge to the graph of $y = x^2$, a polynomial curve. Courtesy of Chand John.

Could smooth curves comprising such fractals be chosen based on meaningful geometric aspects of an image, such as vector fields of most rapid changes in pixel intensity, thereby tying biomedically relevant properties of fractals back to other geometric features of biomedical images? While more research is needed to decide whether fractals, alongside deep learning, can join the collection of mathematical gems that propel biomedical image analysis to new heights, it's an intriguing idea to explore. □

The polynomial curves on the left fully represent the fractal on the right. Courtesy of Chand John.

DETAILS

Chand John, PhD, is a senior software engineer and educator.

DIVERSE DISCIPLINES, ONE COMMUNITY
BiomedicalComputation REVIEW
Published by the Mobilize Center, an NIH Big Data to Knowledge Center of Excellence

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

Stanford University
318 Campus Drive
Clark Center Room W352
Stanford, CA 94305-5444

SeeingScience

BY KATHARINE MILLER

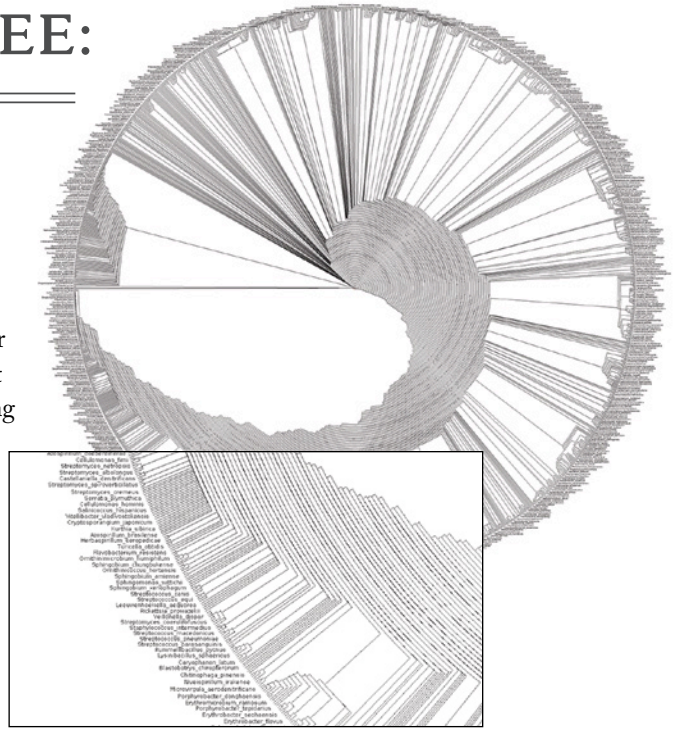
AN AUTOMATED SUPERTREE:

A Model for Extracting Literature-based Knowledge

Too much scientific knowledge is buried in published literature. Case in point: The phylogenetic relationships among microbial species are locked into numerous publications about individual species and their close relatives. And because those publications don't include machine-readable data, the information is difficult to extract. Thus, efforts to create supertrees (large trees assembled from a combination of many smaller phylogenetic trees) typically involve a handful of graduate students doing a massive cut-and-paste job—connecting trees bracket by bracket on a computer. "It's mind-numbingly tedious," says **Ross Mounce, PhD**, Open Access Grants Manager at the Arcadia Fund.

As a postdoctoral research associate at Cambridge University, Mounce set out to create a microbial supertree by using computer vision to extract information from smaller phylogenetic trees in a single journal. "We point the program at the file and it will do its best to extract phylogenetic data from the image," Mounce says. The result is not the best tree, Mounce says, but a proof-of-concept for developing a scalable, automated process. "It's a solvable problem," he says, that has been made easier in the United Kingdom by recent changes to copyright laws. As long as a researcher has legitimate access to a published piece of literature (through a university library, for example), "it's legal to do sophisticated analyses on it without asking permission of the copyright holder," Mounce says. Without that legal right, it would be nearly impossible to perform scalable syntheses of the literature.

"The future is really exciting, because if you had an ongoing reproducible pipeline, you could have a tree of life that self-updates every day," Mounce says. The same is true for any piece of scientific knowledge: "You could check back and see a self-updating synthesis of the current evidence on any topic," he says. "That's the idea really." □



Using an automated, scalable method, Mounce and his colleagues applied computer vision techniques to automatically convert phylogenetic trees from figures in a single journal (the International Journal of Systematics and Evolutionary Microbiology*) back into re-usable, computable, phylogenetic data. They then used established supertree methods to generate the tree of microbial life shown here. Reprinted from Mounce R, Murray-Rust P, Wills M, A machine-compiled microbial supertree from figure-mining thousands of papers.* Research Ideas and Outcomes *3: e13589. https://doi.org/10.3897/rio.3.e13589, (2017).*