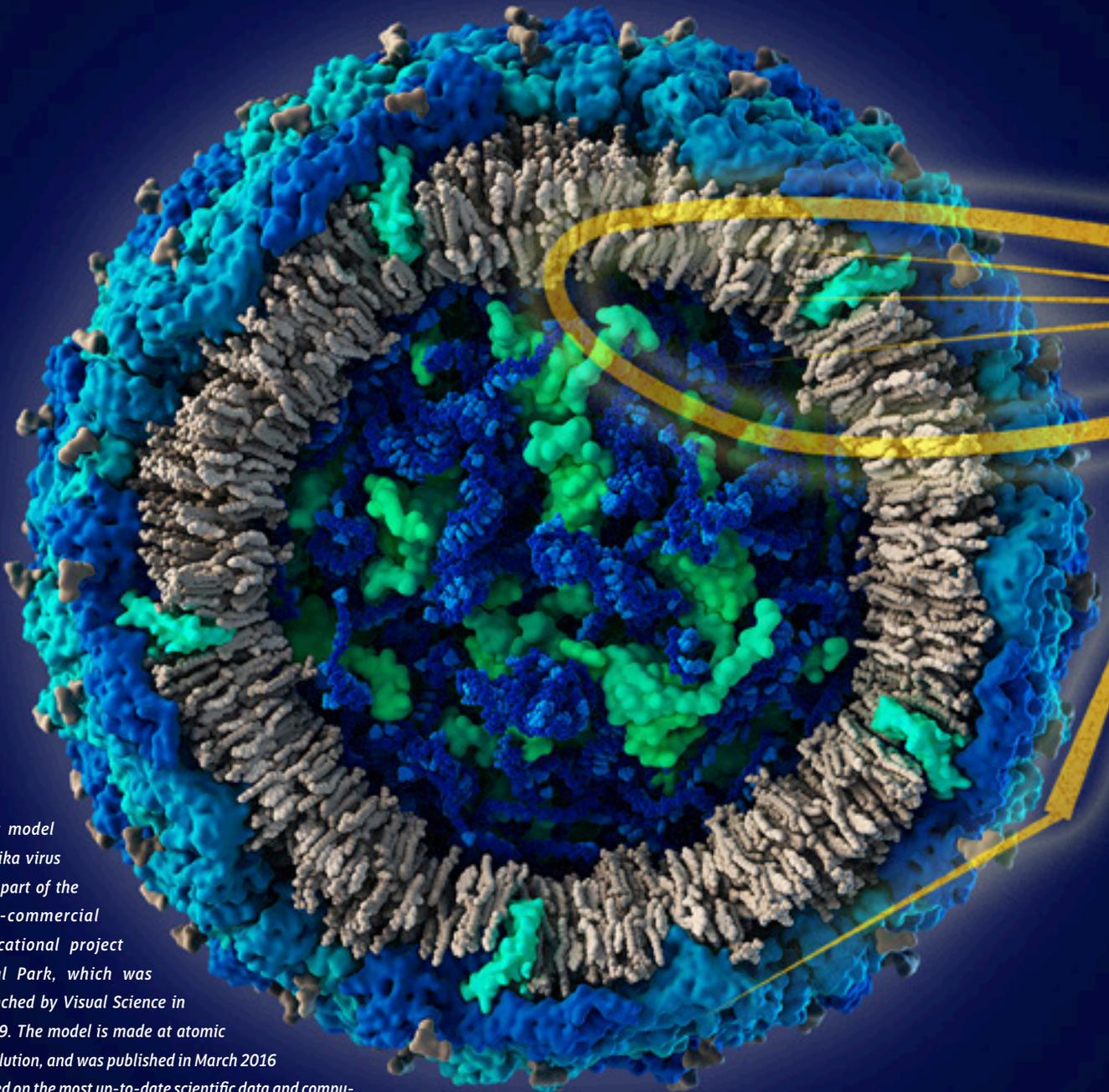


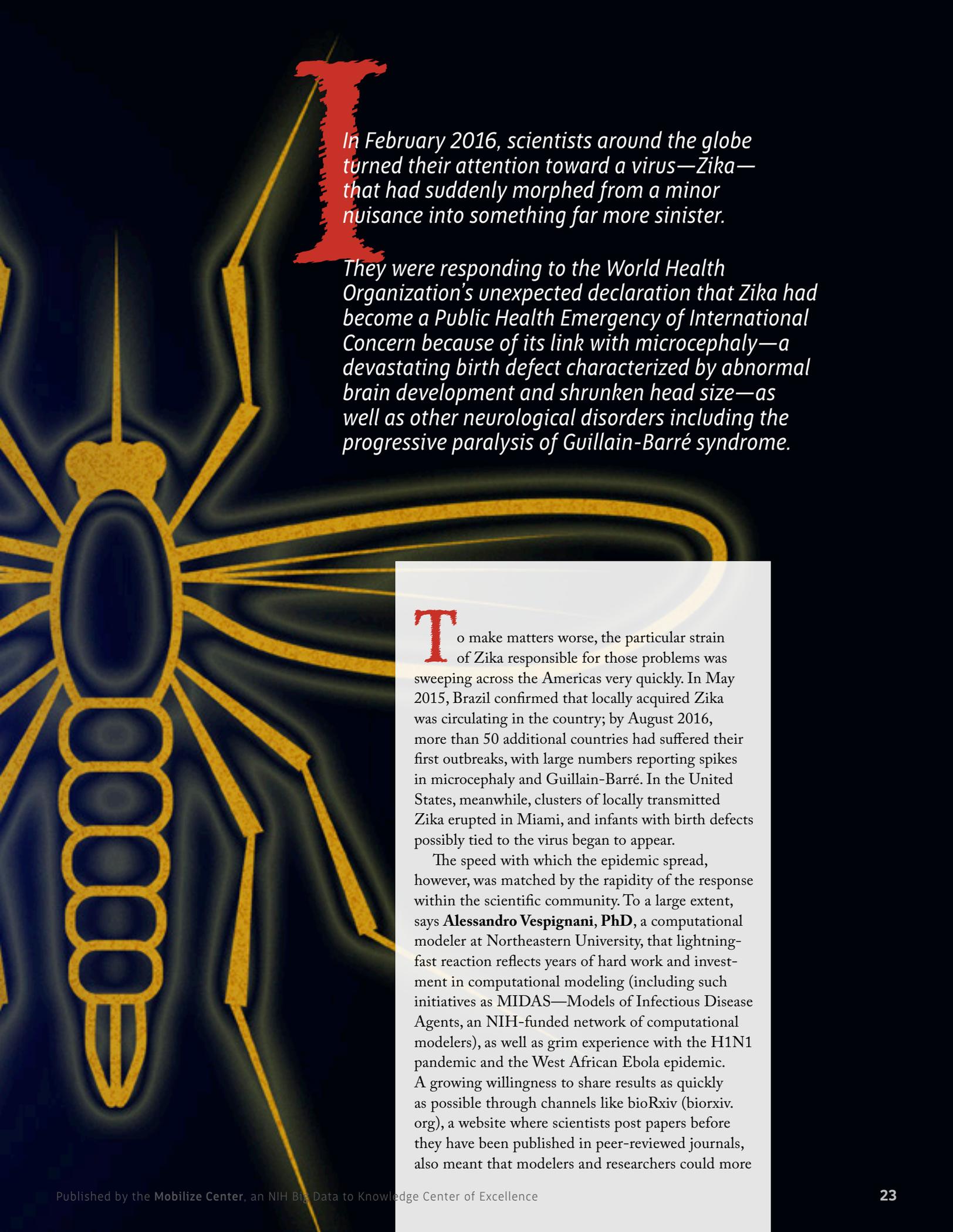
# Zika!

BY ALEXANDER GELFAND

## Computational Biology to the Rescue



*This model of Zika virus is a part of the non-commercial educational project Viral Park, which was launched by Visual Science in 2009. The model is made at atomic resolution, and was published in March 2016 based on the most up-to-date scientific data and computational biology simulations available at that time. Reprinted with permission from Visual Science and creator Ivan Konstantinov. The original image and animation may be found at <http://visual-science.com/projects/zika/3d-model/>*



**I**n February 2016, scientists around the globe turned their attention toward a virus—Zika—that had suddenly morphed from a minor nuisance into something far more sinister.

They were responding to the World Health Organization's unexpected declaration that Zika had become a Public Health Emergency of International Concern because of its link with microcephaly—a devastating birth defect characterized by abnormal brain development and shrunken head size—as well as other neurological disorders including the progressive paralysis of Guillain-Barré syndrome.

**T**o make matters worse, the particular strain of Zika responsible for those problems was sweeping across the Americas very quickly. In May 2015, Brazil confirmed that locally acquired Zika was circulating in the country; by August 2016, more than 50 additional countries had suffered their first outbreaks, with large numbers reporting spikes in microcephaly and Guillain-Barré. In the United States, meanwhile, clusters of locally transmitted Zika erupted in Miami, and infants with birth defects possibly tied to the virus began to appear.

The speed with which the epidemic spread, however, was matched by the rapidity of the response within the scientific community. To a large extent, says **Alessandro Vespignani, PhD**, a computational modeler at Northeastern University, that lightning-fast reaction reflects years of hard work and investment in computational modeling (including such initiatives as MIDAS—Models of Infectious Disease Agents, an NIH-funded network of computational modelers), as well as grim experience with the H1N1 pandemic and the West African Ebola epidemic. A growing willingness to share results as quickly as possible through channels like bioRxiv (biorxiv.org), a website where scientists post papers before they have been published in peer-reviewed journals, also meant that modelers and researchers could more

speedily adapt their computational tools as new information became available.

## First Approximations

Alex Perkins, PhD, a researcher at the University of Notre Dame who studies the dynamics of infectious disease transmission and control, was one of the first modelers to swing into action. His goal: to try to quickly predict

for exploring all of those questions.

Unlike diseases such as Ebola or influenza, which are passed directly from person to person, Zika is a vector-borne disease that is primarily transmitted through the bite of the *Aedes aegypti* mosquito, an insect that thrives in the tropics and whose range and numbers are closely determined by climate. It can also be transmitted sexually, and by another

acquire the virus, then transmit it to uninfected people). But it does provide an opportunity to incorporate finely grained demographic and climate data, including details such as local population and birthrate; average daily temperatures, which govern where and how long the mosquitoes can live and, therefore, how many people they can infect; and even income levels.

---

**Perkins' model, which he first described in a paper posted to bioRxiv less than 2 weeks after the WHO declared a public health emergency, predicts that 93.4 million people, including 1.65 million childbearing women, could be infected before the first wave of the epidemic comes to an end.**

---

the course of the Zika epidemic and the likely number of victims both at home and abroad—information that governments and public health officials could use to plan interventions.

Perkins had for some time been contemplating the problem of integrating disease data collected at different scales. On the one hand, richly detailed local data related to such things as population and climate helps researchers understand factors affecting disease transmission. On the other hand, case reports—i.e., the number of suspected and confirmed cases tallied by hospitals—tend to be collected at the state- or country-wide level.

What, Perkins wondered, might be accomplished if the first bucket of data could be related to the second? And how could scientists predict the course of an epidemic before significant amounts of case data had begun to accumulate? Could they perhaps anticipate the total number of people who might be affected in particular geographic locales while early interventions could still have the greatest possible impact?

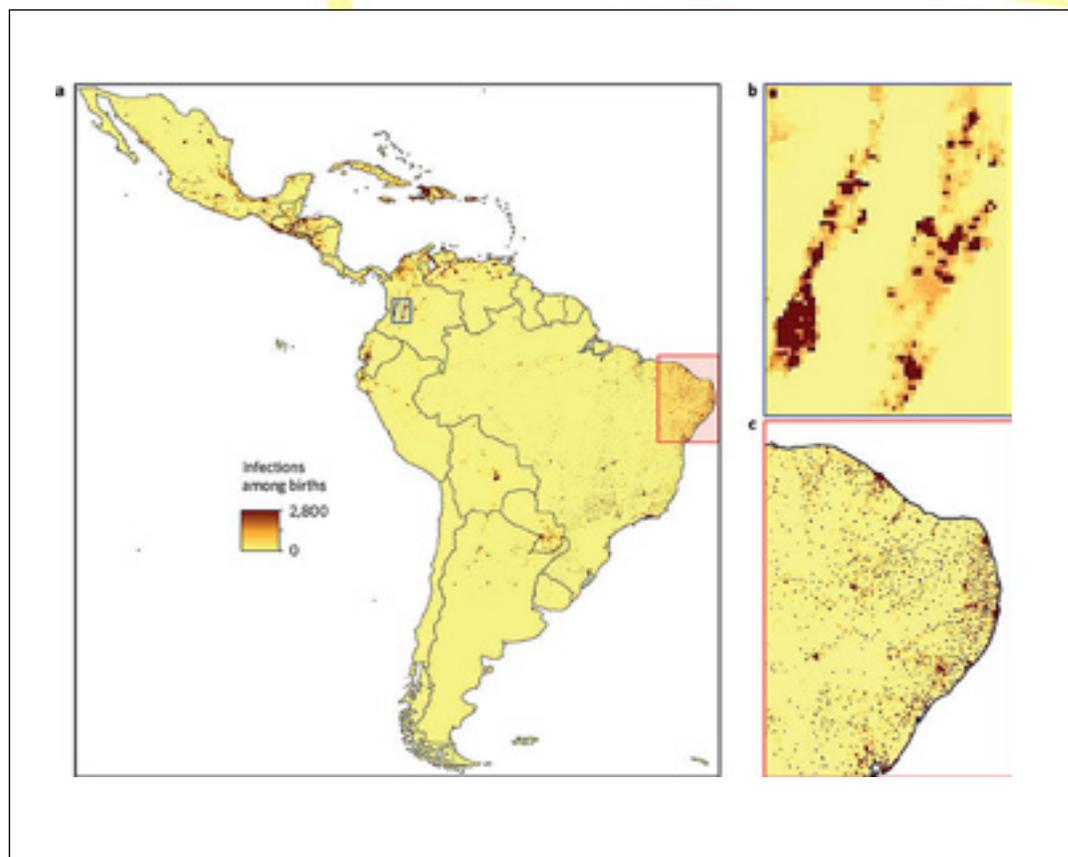
Zika offered both urgent and fertile ground

species of mosquito called *Ae. albopictus*.

This makes Zika difficult to model, since there is more than one infected species to deal with and the chain of transmission is complicated (uninfected mosquitoes bite infected people,

(Affluent people enjoy air conditioning and window screens, which reduce exposure to mosquitoes. Poor people do not, and are therefore at greater risk.)

In addition, while it was first identified more than 60 years ago, Zika remained



Gridded spatial projections of median infections among childbearing women at 5x5 km resolution across Latin America and the Caribbean as a whole (a) and in two specific areas: Cali, Colombia (b) and Recife, Brazil (c). Each grid cell is shaded according to the median number of infections for that cell based on 1,000 simulations. Reprinted by permission from Macmillan Publishers Ltd: TA Perkins, AS Siraj, CW Ruktanonchai, MUG Kraemer, AJ Tatem, Model-based projections of Zika virus infections in childbearing women in the Americas, *Nature Microbiology* 1, Article 16126 (2016).

for most of that time a neglected tropical disease by reason of its mild symptoms (low fever, rash) and lack of known complications. Consequently, when the World Health Organization (WHO) declared a global public health emergency, the scientific community confronted a disease about which it knew remarkably little. Like many, Perkins therefore had to rely on data that had already been collected for other mosquito-borne diseases such as dengue and chikungunya.

Dengue is a member of the same family of viruses as Zika, and all three illnesses are transmitted by *Ae. aegypti*. So Perkins used some of the basic transmission parameters that had already been established for dengue, and looked to a series of previous chikungunya epidemics to get an idea of the total number of people who might be infected. He then used that information, along with mosquito distribution maps and the aforementioned local data, to build a model that could capture the intersection between where people lived, where the mosquitoes were likely to be, and where the conditions were most suitable for transmission—a model that could project the number of infections among the general population, and among pregnant women in particular, across Latin America and the Caribbean, at a resolution of 5km by 5km.

Perkins is quick to note that his model is static rather than dynamic, and therefore estimates only how many people could become infected and not how long that might take. (A less geographically precise dynamic model developed by **Neil Ferguson, PhD**, at Imperial College London does provide such a timeline, predicting, for example, that the current epidemic will burn itself out within three years.) Moreover, it works best at the level of cities, but probably overestimates the total count at the country-wide level.

Even with those caveats in mind, the numbers are eye-popping: Perkins' model, which he first described in a paper posted to bioRxiv less than two weeks after the WHO declared a public health emergency, predicts that 93.4 million people, including 1.65 million childbearing women,

could be infected before the first wave of the epidemic comes to an end.

## **Adventures in the Fourth Dimension**

Perkins himself says that the dynamic model developed by Vespignani at Northeastern offers the best of both worlds: geographically specific estimates of how many people could be infected, and at what speed.

Ironically, when the NIH-funded Center for Inference and Dynamics of Infectious Diseases initially invited Vespignani, who has previously modeled Ebola and the H1N1 virus, to try his hand at Zika, his first reaction was an emphatic no. The reason: He didn't want to have to deal with the mosquitoes.

Vespignani and his colleagues simulate the spread of epidemics across time and space using the Global Epidemic and Mobility Model (GLEAM), a stochastic modeling platform that randomly moves simulated populations of individuals through a series of epidemiological states (susceptible, infected, recovered), generating ensembles of possible scenarios from which the most likely future path of an epidemic can be estimated. It even takes into account the way people travel from place to place, spreading disease as they go.

That already adds up to a lot of complexity, even for diseases that are transmitted directly between people. Throw in a couple of vectors like *Ae. aegypti* and *Ae. albopictus*, which can't travel very far (a typical mosquito only flies an average of 400 meters in its lifetime), and you suddenly have to simulate a whole new population of disease-bearing individuals and their movements at a very high level of detail—individuals whose range and lifespan depend heavily on temperature, and may therefore change drastically from season to season. (Mosquitoes die more quickly in winter than in summer, and if their lifespan drops below Zika's incubation period, they cannot transmit the virus at all.)

Vespignani initially assumed that achieving that level of detail in GLEAM would be impossible, and only changed his mind when he saw the rich

mosquito-related data that vector biologists at the Centers for Disease Control (CDC) and elsewhere had pulled together. "It was really a learning experience," he says, adding that having expanded GLEAM to accommodate one vector-borne disease, he and his collaborators should now be able to simulate others.

Vespignani and his team took into account many of the same factors (e.g., mosquito distribution, wealth) that Perkins' model used. Because GLEAM is able to simulate the course of an epidemic over time, however, Vespignani asked somewhat different questions: What would the timeline of the Zika outbreak look like from place to place? What would its impact look like at specific points in time? And when, exactly, did the virus first arrive in Brazil?

For a modeler, that last question is important, since the reliability of a model's projections depends on its ability to reconstruct the past. "You must get the past right in order to get the future

---

**GLEAM [the approach used by Vespignani's team] consistently predicted a slow-moving epidemic that would manifest in multiple waves in some places (Honduras, Mexico, Puerto Rico) due to seasonal effects.**

---

right," Vespignani says. He was therefore reassured when GLEAM determined that Zika was most likely introduced to Brazil in 2013, a finding that agreed with the results of phylogenetic and molecular clock analyses performed by **Oliver Pybus, PhD**, of Oxford University, and colleagues in Brazil.

Yet even with a well-calibrated model, forecasting the course of the epidemic was not straightforward. Like Perkins, for example, Vespignani had to cadge some of his transmission parameters from dengue, introducing a degree of uncertainty into his calculations. Because Zika is

passed from humans to mosquitoes and back again, there is also some fuzziness surrounding the serial interval, or the time between one infection and the next. And no one really knows how much of a role *Ae. albopictus* plays in spreading the disease. As a result, Vespignani performed several rounds of sensitivity analysis, essentially playing with small variations in

of infections across the U.S. on a state-by-state basis, a task that requires performing millions of simulations on 30,000 processors on a cloud computing platform. Their efforts generated headlines when GLEAM estimated that there could be 25 times the number of travel-related cases reported by the CDC, but Vespignani says that wasn't really

on the ground through diagnosis and treatment is another. Yet here, too, computation is playing an important role.

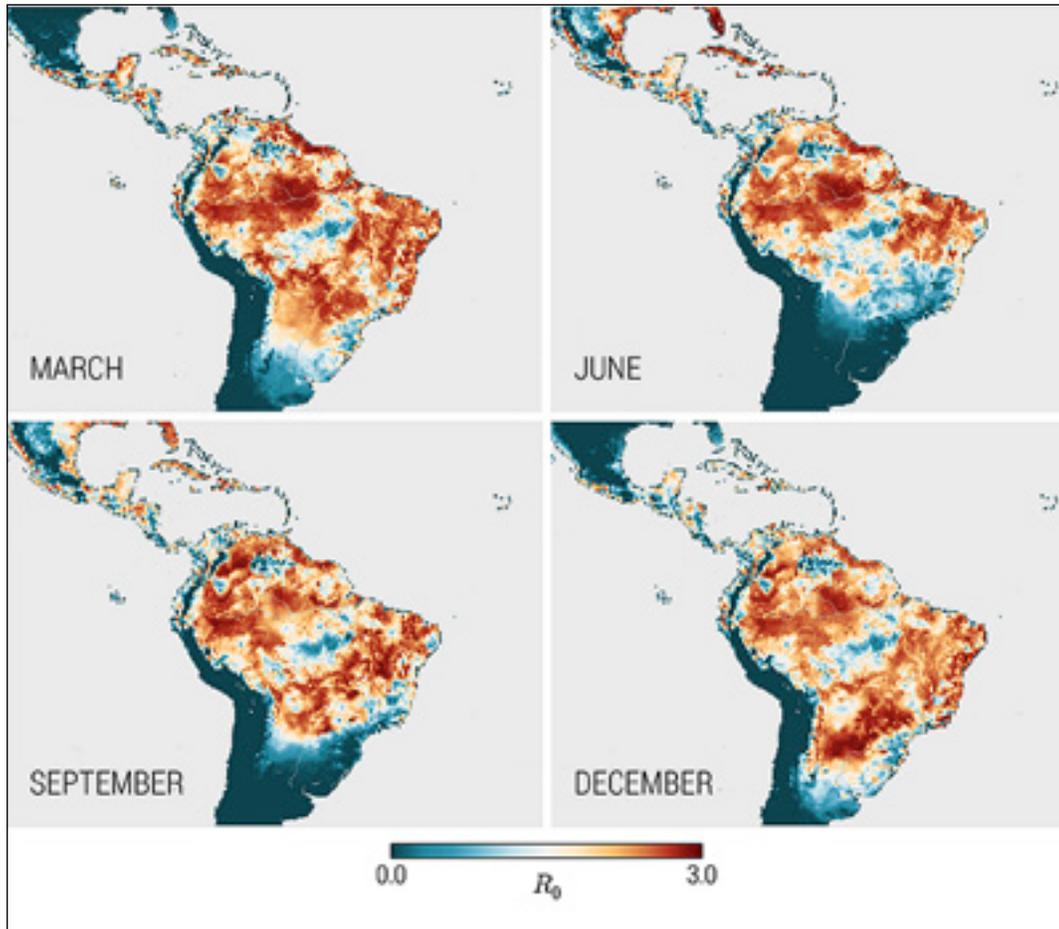
Standard diagnostic methods such as antibody detection aren't ideal for Zika because false positives can arise among people who have previously been infected by a related virus such as dengue. But the cost and complexity of more accurate methods such as DNA or RNA detection puts them beyond the reach of basic health clinics in poor, remote areas.

Now, however, a team of scientists assembled by **James J. Collins, PhD**, of MIT and Harvard's Wyss Institute, is changing that. Together, they have created a cheap, quick, and highly sensitive RNA test that could be used practically anywhere.

Originally developed to detect Ebola, the test relies on two pieces of technology: programmable RNA sensors called toehold switches that can be designed to detect virtually any RNA sequence; and a freeze-dried, paper-based platform that allows those toehold switches to be stored at room temperature on little paper discs, and activated simply by adding a bit of blood plasma and some water.

The switches are made of synthetic strands of RNA that encode a reporter protein that can make the paper change color from yellow to purple. But the switches also contain a hairpin structure called a stem that physically prevents the RNA from being translated unless the stem itself is unwound. In an ingenious twist, the switches are also configured to be perfectly complementary to specific target sequences of RNA. Only when the switches encounter their targets do their stems unwind, allowing the reporter protein to be produced and causing the paper to change color. Diagnosis: positive.

**Alexander Green, PhD**, who developed the switches as a postdoctoral fellow at the Wyss and is now on faculty



Monthly seasonality for the time- and location-dependent basic reproduction number—the number of people that each infected individual is expected to infect. The Equatorial region presents less seasonality than the non-Equatorial regions, where the changes of the season have a strong impact over the temperature and consequently over the basic reproduction number. Reprinted from Q Zhang, K Sun, M Chinazzi, et al., *Projected spread of Zika virus in the Americas*, <http://dx.doi.org/10.1101/066456>.

parameters—changing the serial interval, for example, or removing *Ae. albopictus* from the picture altogether—to see if the model would break down. (The various scenarios can be seen at [zika-model.org](http://zika-model.org).)

It didn't. Instead, GLEAM consistently predicted a slow-moving epidemic that would manifest in multiple waves in some places (Honduras, Mexico, Puerto Rico) due to seasonal effects.

Vespignani and his team are currently projecting the total possible number

surprising: Only 20 percent of infected people show symptoms, and those tend to be so mild that Vespignani himself doubts that he would go to the hospital if he had them. More reassuringly, the model predicts that this country will only see relatively small outbreaks of the sort that have already occurred in Florida.

### A Quick and Easy Test

Anticipating the course of an epidemic is one thing; dealing with it

at Arizona State University, explains that computation is involved at several levels.

For one thing, in order for the sensors to detect the minute quantities of Zika

---

## A team of scientists assembled by James J. Collins, PhD, of MIT and Harvard's Wyss Institute, ... have created a cheap, quick, and highly sensitive RNA test [of Zika infection] that could be used practically anywhere.

---

RNA present in the blood of an infected person, those target sequences must first be amplified. Yet amplification itself makes use of short nucleotide sequences called primers, and if those aren't chosen wisely, trouble may ensue.

If the primers aren't specific enough, for instance, they may also amplify other, similar sequences, like those belonging to dengue.

For another, not all RNA sequences are equally well-suited to detection by toehold switches. When a switch meets its target, the two strands of RNA intertwine, their bases binding to one another; and that interaction can interfere with the performance of the switch itself. Not all switches are equally sturdy, either; and if the stem is too weak, it might unwind even in the absence of target RNA.

Green and his colleagues therefore used several different algorithms and software tools—some custom-built, others open-source—to rationalize both primer selection and switch design.

First, they used their toolkit to screen the Zika genome for regions that were compatible with RNA amplification, filtering out those that were too similar to closely related

viruses such as dengue or to human RNA. With a list of candidate target sequences in hand, they then simulated every toehold switch that could conceivably

bind to those potential targets, and evaluated which combinations of primer and switch would work best.

It took less than a day to construct and test the computationally optimized switches, which were sensitive enough to detect Zika in blood plasma samples and specific enough not to be fooled by dengue. And manufacturing a disk of freeze-

dried paper loaded with switches and amplification materials costs only a dollar.

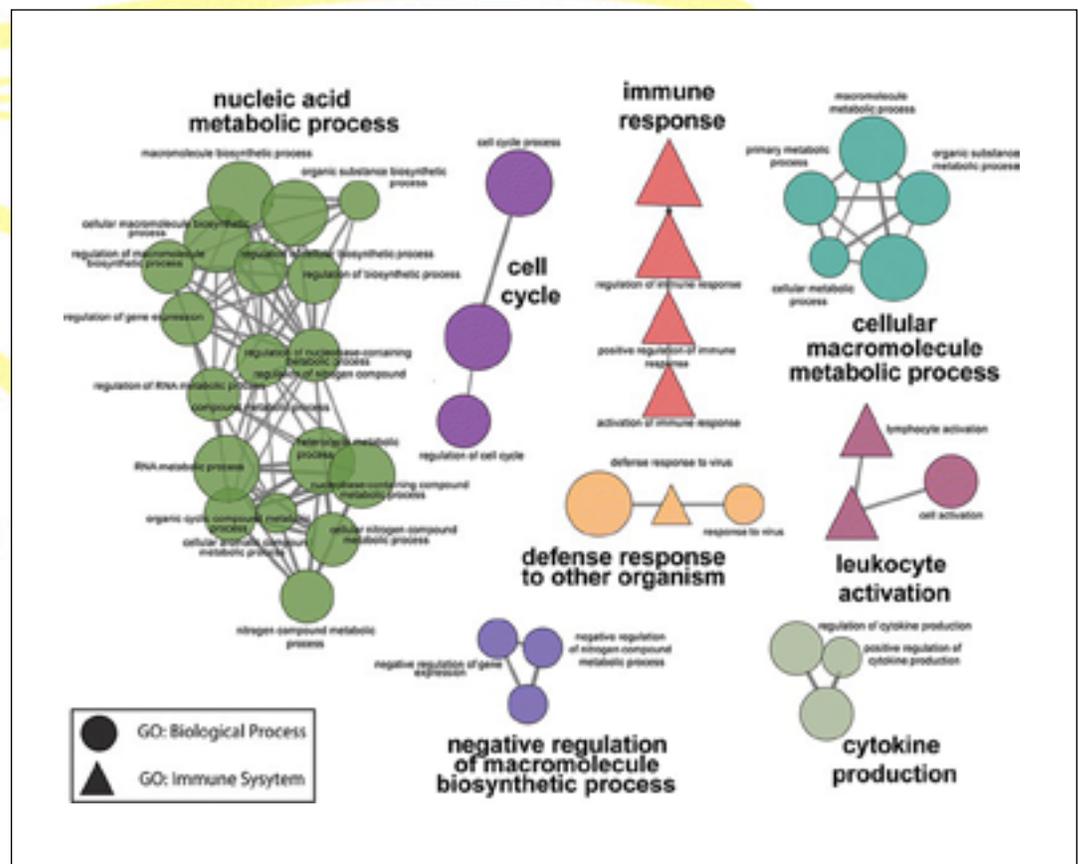
Greene and his colleagues hope to make the test even quicker and less expensive. They also plan to validate

their system using human samples, and to extend its range so that it can detect other pathogens as well.

## Mapping the Mechanisms of Disease

Of course, once you've diagnosed a disease, the next step is treating it. Which is why researchers are also trying to understand exactly how Zika causes microcephaly and other neurological disorders, and are working to find drugs that can fight it.

**Yi Ren, PhD**, a cell biologist at Florida State University who studies inflammation, is one of those trying to get a handle on how Zika does its damage. Her FSU colleague **Hengli Tang, PhD**, was among the first to explain how Zika could cause microcephaly in fetuses—namely, by



*When human neural stem cells are infected with the Zika virus, certain genes are over- and under-expressed compared with normal cells. This visual map shows the various networks of biological processes and immune system responses related to those over- and under-expressed genes. These networks were generated using the Gene Ontology (GO) database and the open-source software platform Cytoscape. Circles represent biological processes (e.g., metabolic processes) in the Gene Ontology database, while triangles represent immune system responses. Groupings with less than three connections were excluded from the final list of networks. Reprinted from AJ Rolfe, DB Bosco, J Wang et al., *Bioinformatic analysis reveals the expression of unique transcriptomic signatures in Zika virus infected human neural stem cells*, Cell & Bioscience 6:42 (2016).*

disrupting cell division and causing death among the neural progenitor cells that give rise to the various components of the nervous system—and Ren, in turn, wondered what inflammatory pathways the virus might activate.

**Alyssa Rolfe**, a PhD student in Ren's lab, explains that she and her colleagues used a variety of bioinformatic tools to analyze the RNA sequence data from Tang's Zika-infected human neural progenitor cells (hNPCs) in order to learn more about how the virus does its dirty work—and to suggest potential strategies for thwarting it.

After assembling a list of all of the genes that were either over-expressed or under-expressed in Tang's Zika-infected cells, the team used the Gene Ontology database to figure out which basic cellular functions those differentially expressed genes might be affecting. They also compared their list with the genes associated with six different neurological diseases in MalaCards, a searchable database of human diseases and disorders; and used an open-source software platform called Cytoscape to create a visual map of all the networks of intracellular biological processes and immune system responses associated with the up- and down-regulated genes. They even compared the gene expression profile of the Zika-infected cells to the profile of hNPCs that were infected with cytomegalovirus (CMV), which can cause a battery of birth defects including microcephaly.

The results were intriguing and, at times, unexpected. For example, the MalaCards search indicated that the pattern of gene expression in the Zika-infected cells had more in common with a suite of congenital nervous system disorders than it did with Guillain-Barré syndrome. And there was little correlation between the immune response pathways that were up- or down-regulated in the Zika-infected cells and their CMV-infected counterparts, suggesting that while the two viruses can cause comparable birth defects, they do so through different mechanisms. Moreover, four of the eight networks the team identified through

visual mapping were associated with immune responses—a surprise, says Rolfe, since one wouldn't expect hNPCs to have any significant interaction with the immune system at all.

The real shocker, however, came when the team dug deeper into the immune and inflammatory pathways associated with their list of genes. Rolfe and her colleagues discovered that a number of genes that one would only expect to see expressed in various kinds of immune cells, such as T-cells and dendritic cells, were in fact over- and under-expressed in the infected neural progenitor cells. “You wouldn't think those genes would have any function in hNPCs,” Rolfe says.

It's possible, she explains, that Zika is either pushing those cells to differentiate into some unknown state; or that the virus is somehow encouraging hNPCs, which do have an innate capacity to modify or regulate immune functions—producing proteins called cytokines, for instance, that normally promote healthy neural development—to shift from an anti-inflammatory role to a pro-inflammatory one. The second possibility, in particular, raises the question of what that shift might do to a developing fetus, and whether moderating the resulting inflammation might limit the negative consequences of infection.

Rolfe says that further investigation in a wet lab will be necessary to sort all of that out. But she hopes that the bioinformatic analysis she and her colleagues have already done will give other researchers useful clues for mitigating Zika's impact.

### **Going Viral on the Grid**

While Rolfe and the rest of Ren's team are probing for insights that could lead to fresh strategies for fighting Zika and its terrible effects, the researchers behind OpenZika ([openzika.ufg.br](http://openzika.ufg.br)) are using computation to virtually screen millions of existing compounds for ones that might already do the trick. The idea, explains **Joel S. Freundlich, PhD**, a chemist at Rutgers New Jersey Medical School who is collaborating on the project, is to jumpstart drug discovery by computationally whittling down the massive list of possible drug candidates

to a more manageable set of likely prospects that can be tested in the lab.

Given the numbers involved, that winnowing process is important. **Alexander L. Perryman, PhD**, a co-principal investigator on the project who works as a senior researcher in Freundlich's lab, points out that even using high-throughput methods, most laboratories can only screen a couple thousand to a few hundred thousand compounds at a go, with Big Pharma pumping that number up to “a couple of million.” The OpenZika team, on the other hand, is screening 8000 FDA- and EU-approved drugs and NIH drug candidates, plus another 6 million compounds pooled from various sources to see if any are likely to disable or kill the Zika virus, with an additional 38 million compounds waiting in the wings.

OpenZika performs virtual experiments known as docking calculations that predict how small, drug-like molecules will bind and interact with the proteins that scientists suspect allow Zika to infect its victims and replicate inside them. And it does so on IBM's World Community Grid (WCG), which draws its computational horsepower from more than 700,000 volunteers in 80 countries who donate processing time on their idle computers, smart phones, and tablets, creating what Perryman calls “one of the largest supercomputers on the planet.” (Perryman previously used WCG to drive computational drug discovery projects for malaria and HIV/AIDS.)

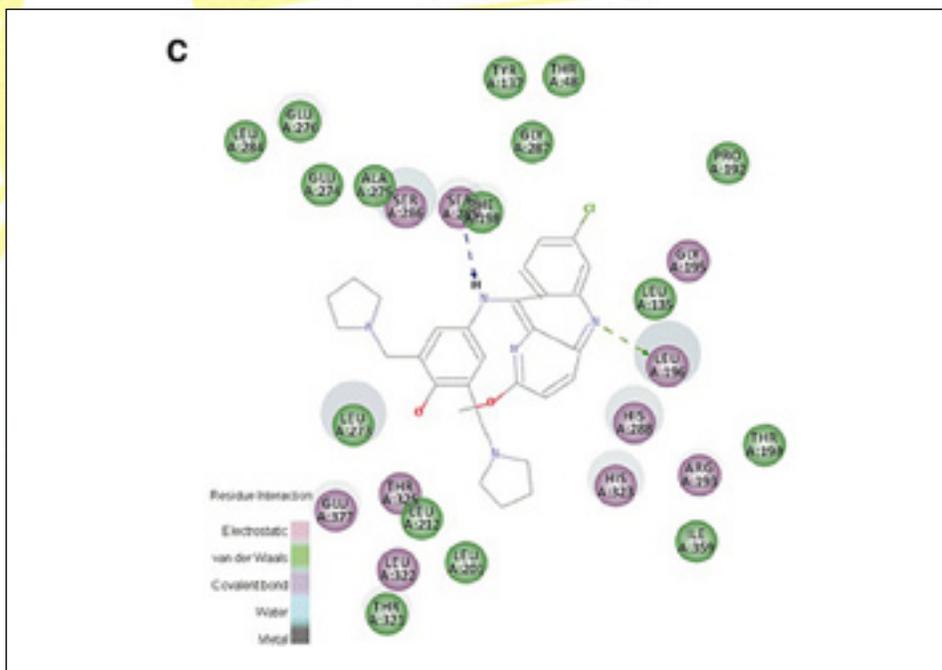
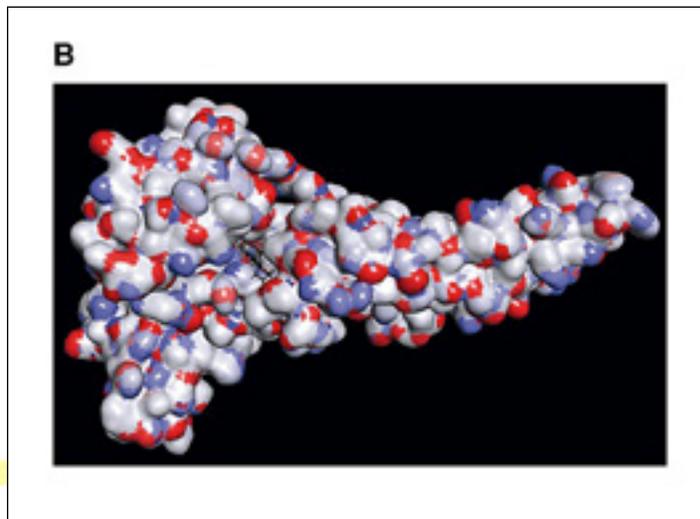
The team employs a program called AutoDock Vina to predict the interactions between the small molecules in its compound libraries and various Zika proteins, virtually “docking” flexible 3-D atomic-scale models of the former to the latter in hopes of identifying molecules that can inhibit the virus's ability to function.

Each virtual experiment, or docking job, calculates the interactions between a single binding site on one protein, and one small molecule that is placed in a variety of positions, conformations, and orientations. That adds up to a lot of calculations. But WCG can handle it: Whereas most researchers

who use supercomputers measure their allotted processing time in thousands of CPU hours, the OpenZika team counts its share in thousands of CPU years. Within the first three months of the project, Perryman had submitted approximately 900 million docking jobs and received 439 million results.

viruses such as dengue and yellow fever. As scientists began to generate structures for the Zika proteins themselves, the OpenZika researchers incorporated those as well; but they continue to use data from related viruses in part because they hope to find broad-spectrum antivirals that will work against more than one.

example, AutoDock Vina thinned the herd to 160; Perryman trimmed it to 15; and Freundlich and Ekins used their chemical expertise to eliminate all but 8. (Eventually, they plan to use Bayesian machine-learning algorithms to do more of that filtering for them.) Of those, five will be sent to



*The OpenZika team used freely available online software to create a homology model of the ZIKV envelope protein (A) that could then be used to dock compounds (such as pyronaridine, shown docked in B and C) and score them in order to prioritize compounds for in vitro testing. Reprinted from S Ekins, D Mietchen, M Coffee, et al. Open drug discovery for the Zika virus. F1000Research 2016, 5:150 (doi: 10.12688/f1000research.8013.1)*

Because no one had bothered to determine the physical structure of the various components of the Zika virus before the current epidemic began, Perryman and his colleagues initially had to rely on speculative 3-D computational renderings, or homology models, of the Zika proteins that various team members created using the Zika genome and structural data gleaned from related

The software scores the performance of each compound, estimating the likelihood that it will stop Zika in its tracks. After that, the humans step in, visually inspecting the highest-scoring compounds to determine which might be the best drug candidates. Of the 8,000 drugs and drug candidates that have been screened against one particularly promising target, for

collaborators at the University of California, San Diego, who will run cell-based experiments to see if the compounds really are as good as they seem. If they are, the medicinal chemists on the team will try to determine what makes them effective so that they can be made even more potent, even as the team continues to screen its compound libraries against yet another target in order to identify more prospects for lab testing. The goal is to get the most promising candidates into the lab and back out the door in enhanced form as quickly as possible.

Forecasting. Diagnosis. Basic biology and drug discovery. All have a role to play in dealing with what is shaping up to be one of the greatest global public health crises in recent times. And computation, in turn, is playing a key role in all of them. □