

Anne-Ruxandra Carvunis's studies show that translation is widespread, even in the dark corners of our DNA.





CARVUNIS FINDS MEANING IN  
THE GENOME'S GREAT UNKNOWN  
BY ELAINE VITONE

# DNA'S DARK MATTER

**F**or decades, we thought that genes were a lot like us: forged from the same stuff as our parents, and their parents before them, and so on, dating all the way back to Common Ancestor Immemorial. Every gene on Earth was thought to have used as its template one of the small number of genes that were around when life began.

But then, when it became possible to compare the genomes of various species against each other, researchers started finding misfits—so-called “orphan” genes that looked nothing like their neighbors. They didn’t have any counterparts in other species either—not even in close cousins. If we Earthlings all got here solely by gene duplication, this made no sense. So for years, many scientists didn’t believe orphan genes were real genes.



And if you sequenced the genome of an organism and found something that looked like a gene but didn't have a "family"? Sorry, it couldn't be a gene.

In 2006, a group at Harvard University was scratching their heads over how the literature could've been so wrong about their model organism, yeast. Anne-Ruxandra Carvunis, then an aspiring PhD student, joined the lab just as they were taking a closer look at these orphans and finding their behavior remarkably . . . unremarkable.

They were just ordinary genes, albeit oddballs.

Carvunis, who's now an assistant professor of computational and systems biology at the University of Pittsburgh, was perplexed. *Wait. These genes are perfectly normal, but they don't have families? So where do they come from?* To answer this lingering question, Carvunis looked to evolution, which has become a focal point of her career.

"I didn't have a passion for it growing up," she says. "I mean, like a lot of children, I was a fan of dinosaurs and all that, but I didn't think [evolution] was my scientific calling. It just came because of data that pulled me in. And, once you're in, you're in."

Carvunis ended up studying network biology at Harvard for her PhD. In parallel with her dissertation, which was on protein inter-

for the field. Search Google Scholar today, and "*de novo* gene birth" yields 325 hits. More than 250 of them cite Carvunis.

As it turns out, within the genome, there's an awful lot going on beneath the surface.

In humans, the 20,000 protein-coding genes that researchers typically study only account for about 25 percent of our DNA. Then there's "the rest," Carvunis explains—a mysterious expanse that some call "dark matter" or, far less flattering, "junk DNA," basically, because nobody could figure out what it was there for. It's turbulent, constantly changing. It's also very messy: tons of repetition and traces of our bodies' many tangles with viruses along the way. ("We're very virus-y," she says.)

In her studies of brewer's yeast, Carvunis examined 108,000 short sequences from that genome's great unknown and found that more than 1,000 of these elements were engaged with the cell's protein-production machine—evidence that the so-called junk had the potential to become proteinaceous.

For reasons like this, many prefer the term "intergenic" to "junk."

And, amid darkness and chaos, Carvunis saw order. If a new element was bad for the cell, it was game over for that material. If it was neutral, what happened next was left to chance. And if the element turned into something useful, then natural selection could take hold. Beneficial mutations would snowball, and eventually, this little nugget of nothingness would gain all the characteristics of a gene, invented wholly from scratch.

"So, those elements—I call them proto-genes," says Carvunis, "I found thousands

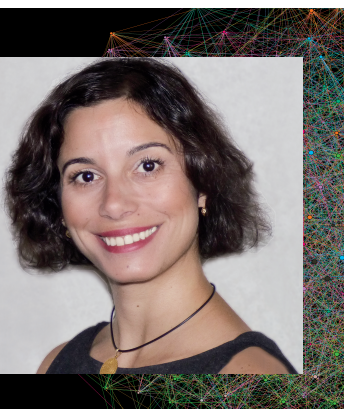
*The New York Times* and *New Scientist*. This spring, she was named a Searle Scholar, one of the most prestigious honors awarded to early career biologists. Her proto-gene paper remains a popular favorite in scientific journal clubs from a variety of fields. At conferences, and in e-mails from young scientists around the world, she often hears that her work has been a source of inspiration—and a reason to rethink dissertations.

Carvunis's colleagues will tell you she loves big ideas. She relishes a good confab over morning coffee in the lab. (The native Parisian spent much of her teen years haunting cafes and talking, talking, talking with friends). She loves to talk science. And broader context/perspective. And evolution! And Where All This Is Going! (Sometimes, when she gets *really* excited, she lapses into French.)

In her latest paper, a coauthored commentary in *Nature Immunology*, Carvunis applied her evolutionary systems biology approach to one of the most perplexing challenges in biomedicine: Why is it that 90 percent of phase I clinical trials fail to advance? That is: Why *isn't* what's good for the mouse more frequently good for the man or woman, as well?

Today, we know that somewhere between 2 percent and 30 percent of genes are *de novo*. Not much is known about all these newly identified genes as of yet, but we do know *de novo* genes in general vary a lot between species. And between individuals (among yeast, anyway), they vary significantly, as well, in terms of their presence, sequence, and expression.

And so Carvunis says it's tempting to



Carvunis

**"So, those elements—I call them proto-genes," says Carvunis,**

**"I found thousands of them in the yeast, which only has 6,000 genes. It was crazy."**

actions, she began to design her own studies of *de novo* genes, as the literature had begun to call them (from the Latin word for "new"). At the time, just over a dozen papers on the topic existed.

After Carvunis finished her doctorate, she wrote up an exhaustive *de novo* gene treatise, using yeast as her model. The paper, which was published in *Nature* in 2012, proposed a plausible mechanism for *de novo*-gene genesis for the first time. And it marked a tipping point

of them in the yeast, which only has 6,000 genes. It was crazy."

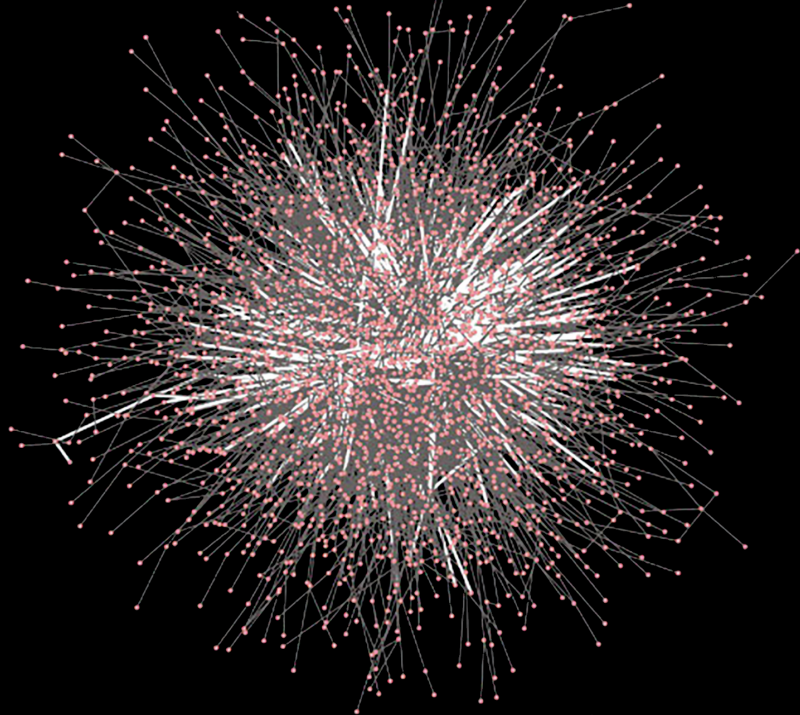
In January 2017, Carvunis came to Pitt as a cofounder and executive committee member of the Pittsburgh Center for Evolutionary Biology and Medicine (slated to open this summer). At 37, she's an international leader in evolutionary systems biology—a new field at the nexus of evolutionary theory, genomics, and computational and systems biology. She's been quoted in stories about gene birth in

explore: Could *de novo* genes have implications for what makes humans, humans, and mice, mice? Or, what separates sickness from health?

"We *don't* know," she says, sounding delighted by this open end.

Carvunis's work isn't focused on one gene, or one specific question of how something works. "We are trying to understand, generally, the whole genome, the whole cell—actually, the whole organism," she





As a PhD student, Carvunis built the first-ever interaction map for a plant's proteins. The figures here, from a 2011 *Science* paper, show these interactions are not organized randomly. If they were, the network would look like a big meaningless hairball (see simulation, top image). The bottom image shows what Carvunis found: communities of proteins that work closely with one another on specific biological processes. And in another *Science* paper published that very same day, Carvunis described how the job of one of these communities is to fight off different types of pathogens. The proteins in this community evolve rapidly to keep up an arms race with their microbial enemies.



says. “How does it work from a systems point of view, or really, from an evolutionary point of view?”

*De novo* gene birth is still a long way off from scientific consensus. Some people tell her they find her work very novel. Others sort of nod and agree, *Well, yeah, it's obvious!* And then there are holdouts who still cry “orphan” and reject this field altogether.

She laughs. “That’s how I know it’s really interesting and worth pursuing.”

In the 17th century, English scientist Robert Hooke magnified a thin sliver of cork and spied thousands of hollow box shapes. To him, they looked like the spare little quarters where monks lived. So he dubbed them “cells.”

Later, industrial era scientists reimagined these fundamental elements of life in the likenesses of engines, boats, and lenticular bridges, the mechanical achievements of their time.

And at the start of the 21st century, scientists of the Internet age looked to the natural

world and saw networks abound—in food webs, for example, and certainly in the inner workings of cells.

“Like any model of the world, our view of the cell is inescapably bound by the time and place in which we live,” Carvunis wrote in a paper published during her postdoc. Scientists, she believes, don’t just automatically shed their perspectives when they clock in at work. (Put a pin in that paper—we’ll come back to it.)

Carvunis came of age in the network years; she was just starting college in Y2K. At first,



she was drawn to neuroscience, intrigued by the complex networks within the brain and curious as to how they might shape consciousness and humanness. “Then I realized that networks are everywhere,” she says.

In the new field of network biology she saw fascinating possibilities: It’s a way to probe both the intricate dynamics of complex systems, and how they amount to so much more than the sum of their parts. It’s quantitative and big picture. It’s collaborative and interdisciplinary. It holds promise in applications to immunology, cancer, developmental biology, and more. And as a computational approach, it could potentially help correct for some of that pesky human subjectiveness in the scientists who apply it. Carvunis was in.

Think of the network biologist’s realm like this:

Our genes write the blueprints for the molecular interactions that keep our cells going. These networks are shaped by their environment (that is, our bodies and *their* environments). And what goes down at that network/environment face-off is exactly why cells are the way they are: sick or healthy, surviving or not.

In other words, networks are the intermediaries between genomes and cells—really, between our genes and us. It’s all the same continuum.

When Carvunis fell hard for evolutionary theory in grad school, things really came together. She began to recognize it pressuring every organism this way and that on a constant basis, shaping our genomes, molecular networks, and phenotypes all at once. (What are phenotypes? Oh, just the differences you can observe within any given species: Tall versus short. Resistant to virus A versus not. Responds to cancer drug X versus doesn’t.)

For her PhD, in the lab of Harvard’s Marc Vidal, Carvunis built the first-ever interaction map for a plant’s proteins. She found that the interactions of proteins made by genes that were products of gene duplications were subject to natural selection. Until then, everyone modeled this as a random process in a constant state of change. She also studied protein interactions between the immune system of plants and the pathogens that plague them and found telltale signs of coevolution—which was also new news.

Both findings made the pages of *Science*.

And she ran both projects concurrently with her groundbreaking *de novo* study.

The proto-gene paper was a big deal, and not just for gene-birth researchers. It threw cellular biology in general for a loop because it challenged a fundamental assumption. We used to think only nonjunk DNA could make a full-fledged protein. Carvunis’s yeast studies showed that, really, translation is widespread, even in the “dark” corners of our DNA.

Also: The proteins that the proto-genomes made? Those were species-specific—and no one had ever even seen them before. Carvunis suspects that these previously unrecognized “proto-peptides,” as she dubbed them, may be in all complex organisms—which would mean a possible treasure trove of new leads for drug discovery, if she’s right.

And so far, her hypothesis is holding up. Just this March, a team in Barcelona showed widespread transcription in the intergenic regions of a mouse (that was in *Nature Ecology and Evolution*).

Mar Albà, the lead author on that paper who heads the Evolutionary Genomics group at the Research Unit on Biomedical Informatics in Barcelona, is one of the few scholars who studied *de novo* genes before they were cool. She says Carvunis’s 2012 paper unified many disparate threads into a cohesive whole and “had a huge impact on the field.”

Albà and Carvunis are collaborators. They got together via e-mail a couple of years ago, along with teams in Germany and Croatia, for a multi-institutional response to a paper out of the University of Michigan that had broadly criticized *de novo* gene research and the validity of these teams’ methods. (Their search algorithm, the Ann Arbor coauthors posited, wasn’t sensitive enough, and introduced bias into evolutionary pattern inferences.) Soon after she arrived at Pitt, Carvunis and her far-flung colleagues published their reply in the very same journal, *Molecular Biology and Evolution*. (After reanalyzing the Michigan team’s data, removing questionable sequences, and even factoring in a false negative rate of up to 15 percent, Carvunis et al. found the tools of the trade were indeed working reliably.)

Albà and Carvunis didn’t get to meet in person until months later, at a conference in Austin. They’d been following each other’s work for years, and fell effortlessly into that deep, specialized, no-explanations-needed kind

of discussion. The insta-friends felt like they went way back.

“It’s very exciting when you can talk at that level with someone,” says Albà. “You’ve been thinking about the same thing for so long. . . . For a scientist, this is not only a job, it’s a life. It’s really hard to stop when you go home!”

## THE TRANSCRIPTION CLOCK

Carvunis’s postdoc mentor, Trey Ideker—who directs the Cancer Cell Map Initiative, the National Resource for Network Biology, and the San Diego Center for Systems Biology at UC San Diego—clearly misses those Carvunis convo sessions terribly. “She’s totally brilliant,” he says, and gushes about what she accomplished while they were labmates.

During that time, Carvunis’s main project was on how transcriptional networks evolve across species. But she saw a big problem: Each breed of researcher (pardon the phrase) was using a different approach—separate toolboxes for fly studies, bird studies, rodent studies, and so on. The longstanding joke among bioinformatics folk, says one *eLife* editor’s commentary on Carvunis and co.’s work, is that these scientists would sooner share a toothbrush than use someone else’s code—so Carvunis “cleaned everyone’s teeth with the same toothbrush.” That is, the team applied one common analysis methodology in studying raw data gathered from a number of species of complex organisms (insects, birds, and mammals—including humans).

Then they found something no one was expecting.

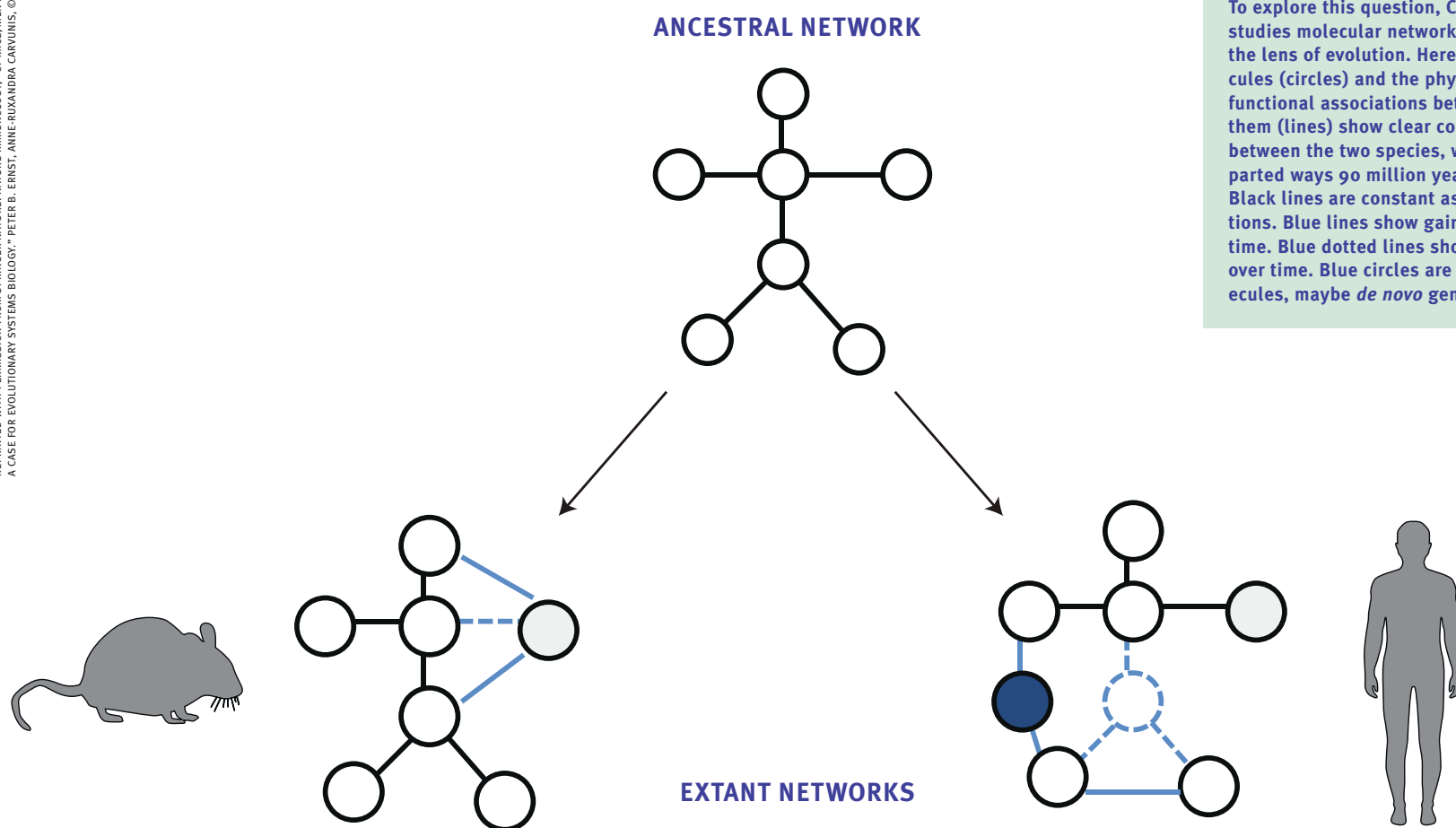
For some reason, even though all these different species reproduce at very different rates, somehow the networks that regulate transcription evolve at the same rate.

In other words, the fly on my wall, for example—whose little fly family will have umpteen generations in the course of my one lifetime—will not evolve any faster than I will. Which raises the question: Do the fly, and I, and the two dozen other species in the team’s sample all have some kind of molecular timekeeper in common?

Carvunis stresses she does not understand what is going on here. Verifying this phenomenon, figuring out its mechanisms (“If your readers have an idea, contact me!”), and probing its consequences are long-term goals



Why *isn't* what's good for the mouse more often good for the human, too? To explore this question, Carvunis studies molecular networks through the lens of evolution. Here, molecules (circles) and the physical or functional associations between them (lines) show clear contrasts between the two species, which parted ways 90 million years ago. Black lines are constant associations. Blue lines show gains over time. Blue dotted lines show losses over time. Blue circles are new molecules, maybe *de novo* genes.



of her lab. But she says it does make sense, if you think about it: “If how species change were really proportional to their reproduction rate, then, from the time I’m a child to the time, I don’t know, I have grandchildren, I cannot tell a story about flies. Because they don’t exist. They’re a completely different animal now.”

In addition to the clock paper, Carvunis and Ideker also collaborated on the beginnings of a project that has since become a major focus in Ideker’s lab. It came out of one of those big-picture coffee talks back in California:

Where, the scientists wondered, was the future of “omics” and high-throughput biology really going? What was the road ahead for big data?

Recall how cells looked like monks’ quarters to the Renaissance scientist, like machinery to the industrialists, and like a network to Generation Web 1.0? Well, naturally, as the San Diego duo pondered this four or five years ago, they first consulted their smartphones.

In 2014, Ideker and Carvunis published

in *Cell* a vision for the way forward in their paper, “Siri of the Cell: What Biology Could Learn from the iPhone.” The new model of the cell, they explained, was turning into much more than data points and straight lines between them. Add your given genome, gene products, metabolites, and other biomolecules, and then link them together with physical interactions and other functional associations, and the schematic gets ... complicated. A bunch of wires running to and fro can’t capture the complexity of multiscaled hierarchies.

But perhaps, eventually, intelligent simulations of cells, tissues, organs, and whole patients could. Imagine a doc asking for a consult: “Hey, Siri? Patient P’s cancer came back—and this time, with new mutations A and B. What should I prescribe now?” It’s early days yet, but Ideker is working to make this dream a reality.

Now, as an independent investigator, Carvunis is building on the blockbuster findings from her graduate work and training. She’s hoping to better understand how proto-genes evolve and create new traits in organisms. Through systematic experiments in yeast,

and complementary computational surveys, she’s working to develop broad taxonomies of proto-gene product functions.

Eventually, she hopes to use these new insights on evolution in an ambitious application, to recreate *de novo* gene birth in the lab—a potential boon for biomedicine that’s hard to even fathom. To do it, her lab is attempting to design a strategy to speed up the evolution of proto-peptides. So far, it’s looking promising, but “this is not even a hint of a submission” to a scientific journal, she’s quick to add.

“It’s very, very exploratory research. But it’s fun,” she says, beaming, and calls this her most exciting project of all.

Nikolaos Vakirlis, a collaborator/mentee from afar, says Carvunis is one of the most enthusiastic people he’s ever worked with. Vakirlis is a postdoc in the Dublin lab of Aoife McLysaght, a founding member of the field of gene birth who identified the first known *de novo* gene in humans. Carvunis and the Irish team are working to identify *de novo* genes in yeast and determine the why, and how, and when of their evolution.



## OF MICE, NOT MEN

Say you're a mouse. You live in a hole in the ground and forage for food among filth. For countless generations, your ilk has been shaped by natural selection for optimal mouse-ness. Your immune system, for example, is the product of millions of years of bottom-feeding and excrement-eating—especially nasty primordial microbes have coevolved with your ancestors' bellies. By now, you are really good at being a mouse.

Now (I know, it's a stretch) let's say you're a human. You are (hopefully) none of the above. Because you and the rodent parted ways, evolutionarily speaking, lo, 90 million years ago.

Animal models of human diseases are just that—models, writes Carvunis in her *Nature*

candidate can even be detrimental to humans, he points out.

In the paper, Carvunis calls it “the illusion of similarity.”

“Mice are not little humans who like cheese,” she tells me in her office in Biomedical Science Tower 3. “We know that. Yet, our genomes are quite similar with mice—80 percent. That's an interesting number. What does it mean? Yes, it's quite similar, but we also know it's the 20 percent that are why we are not mice. So how to identify what matters in those 20 percent?” Some, but not all, will be reasons why you can use a given therapy to cure cancer in a mouse, but not in us. “So can we understand what part of the genome, and the networks, really translate across species,

who would call into question our understanding of what genes are and how they evolve. She would learn that once essence insinuates itself into the scientific literature, it can be hard to extricate. Consider labels like “junk DNA” and “orphan gene.” In yet another collaboration, Carvunis got together with a rhetoric scholar at University of San Diego, a protein chemist in Texas, and a philosopher in Italy. The interdisciplinary team represents a range of personal belief systems, as well, from religious to agnostic. And together they are probing the question of how scholars interpret terminology differently within the sciences—specifically, the term *function*.

In reviewing a sample of the literature, the team found that sometimes scientists wrote

**The fly on my wall, for example—whose little fly family will have umpteen generations in the course of my one lifetime—will not evolve any faster than I will. Which raises the question: Do the fly, and I, and the two dozen other species in the team's sample all have some kind of molecular timekeeper in common?**

*Immunology* commentary. They are approximations. But seen through the lens of evolution, and aided by the tools of network biology, the fundamental differences between these two species aren't just stark; they're concrete.

For this paper, Carvunis teamed up with Peter Ernst, a veterinarian, professor of pathology, and director of comparative pathology and medicine at UC San Diego, who studies animal models of inflammatory bowel disease (IBD). Using several examples of animal-model successes that failed the human test, the team walked through the architecture of how gene products interact with one another, and how that differs between the species.

Let's say, for example, that molecule 1 interacts with molecule 2, which in turn interacts with molecule 3. And that might be the case in both mice and men. But it's not just the molecules that are important—it's how they talk to one another. A bond between molecules in the mouse might be a completely different conversation in you.

“You get enthralled” with the commonalities, says Ernst, “and it can be that the same tissues in both are affected. Or maybe even some of the same cells or molecules. But it ends there.” And sometimes, a given drug

and which cannot? That's my dream.”

This is not, she notes, a way to predict for certain what will and will not work. But it will be a useful tool in ruling out candidates that have very little chance of success.

Ernst notes that, recently, a colleague who was studying inflammatory bowel disease in rodents mapped a genetic region relevant to the disease and found that this region predicted microbiota in people—and their susceptibility to IBD. “And ironically, it's an intergenic region,” Ernst says.

If that pans out in the clinic, he says, “then clearly, that's a new data point that most people would be missing completely.”

## WATCH YOUR LANGUAGE

Jean-Paul Sartre was required reading in Carvunis's high school in Paris. *L'existentialisme est un humanisme* changed her perspective on life, and on science. “Existence precedes essence,” Sartre famously wrote. Individuals create identity and value and meaning—and there is no moral absolute.

Carvunis realized: *What I do is up to me. It's not because somebody said something that it's true. I must go and see and decide for myself.*

It was a fitting launch pad for the scientist

“function” when they really meant “expression.” Other times, they meant “interaction.” And other times they meant what the gene is “there for”—what nature has selected it to do.

The scholars are having so much fun that they're thinking of what to tackle next—perhaps the word *gene*. That one is a huge problem, Carvunis says, because it has so much history; it existed before we even knew about DNA.

“It's very interesting how language and knowledge are intersectionalized,” she says. Scientific knowledge is growing at a fast clip, and language can't keep up. If the words we have don't fit, they can even impede knowledge. “But then we cannot also invent words every five minutes, either.”

Carvunis hopes the team will provide useful frameworks to help scientists become more aware of the words they choose.

And to keep their perspectives in check—they bring them to the bench whether they realize it or not.

“As scientists, sometimes we think that we are just pure minds, but it's not true,” she says. “We are people, and we are inspired by what happens around us.

“We must not forget that. It can be bad, or it can be good. We must make it good.” ■