

# SEARCH

The background of the cover is a complex, abstract digital visualization. It features a dense network of thin, light-colored lines connecting small, glowing nodes, creating a sense of interconnectedness and data flow. Overlaid on this network are several large, semi-transparent rectangular blocks in shades of blue and orange, which appear to be data structures or search results. The overall color palette is dominated by these two colors, with a warm, golden glow emanating from the top right corner, suggesting a bright light source or a high-tech environment.

PERCEIVING THE IMPERCEPTIBLE

FALL 2022 • VOL. 15 • NO. 2 • THE JACKSON LABORATORY

## DEAR FRIENDS,

By the time you read this, it will have been nearly a year since I arrived at The Jackson Laboratory. The months have flown by as I have immersed myself in learning the special attributes that make this complex and dynamic organization unique and powerful. I am more convinced than ever that JAX has a pivotal role to play in transforming biomedical science and human health.

My colleagues and I have been engaged in a strategic planning process that will create a road map to guide us toward our primary goal and opportunity: to translate basic science discoveries into practical advances in the treatment and prevention of disease. We are poised to help improve human health.

As this issue of *Search* highlights, data sciences are increasingly at the core of leading-edge biomedical research. JAX scientists are at the forefront of developing new computational tools and methods, from developing powerful algorithms that can find the significant data points in a vast array of genome sequences, to using machine learning to characterize mouse behavior accurately and on a large scale.

The biggest challenge we confront is not in generating data but in finding effective ways to analyze and synthesize the massive amount of data that we do generate. JAX is positioned to bring together mouse and human data in ways that accelerate the translation of advances in basic biology into new diagnostics and treatments. Our growing investment in data sciences and new technologies complements and amplifies our existing strengths in areas such as rare diseases, neurodegeneration and cancer research.

The Laboratory's newest endowed chair, created by our new Board Chair Tim Dattels, and held by Associate Professor Jennifer Trowbridge, Ph.D., exemplifies our commitment to translation, linking our historic strength in basic science to clinical impact. The establishment of the Dattels Family Chair also illustrates the power of philanthropy to advance our mission. Dr. Trowbridge's lab is changing our understanding of the fundamental genetic and cellular changes that

can lead to blood cancers — and her work is also identifying potential new approaches to treating leukemia that could be game-changing for patients.

Computation alone will not create the path to cures. Basic research in wet labs remains essential. Mouse models and cellular models also have a vital role to play in identifying and testing potential new therapies, before the final step of clinical trials in human patients. And now more than ever, JAX is the place that can bring all of these together.



Lon Cardon, Ph.D., FMedSci  
President and CEO, The Jackson Laboratory

# SEARCH

## ON THE COVER

Modern science produces more data than we can process on our own. JAX researchers are using and creating powerful tools to sift through this data in order to perceive the imperceptible.

**The Jackson Laboratory**  
discovers precise genomic solutions for disease and empowers the global biomedical community in our shared quest to improve human health.

*Search* magazine is produced by JAX Creative.  
Printed November 2022

**EDITOR**  
Danielle Meier

**DESIGN & ART**  
Jane Cha  
Karen Davis  
Danielle Meier  
Zoë Reifsnnyder

**COPY EDITORS**  
Joseph Blanchette  
Carol Lamb  
Rebecca Hope Woods

## FEATURES

- 4 Perceiving the imperceptible
- 6 Beyond observing
- 14 5 questions with Matthew Gerring
- 16 Complex problems, computational solutions
- 18 Putting machine learning to work for cancer patients
- 22 glue: Making sense of the universe in 3D

# PERCEIVING THE IMPERCEPTIBLE

**Our brains are extraordinary.**

BY MARK WANNER

With them, we can acquire input from the world around us, assess current conditions, calculate potential changes and predict future states with astonishing speed. But sometimes, they need help.

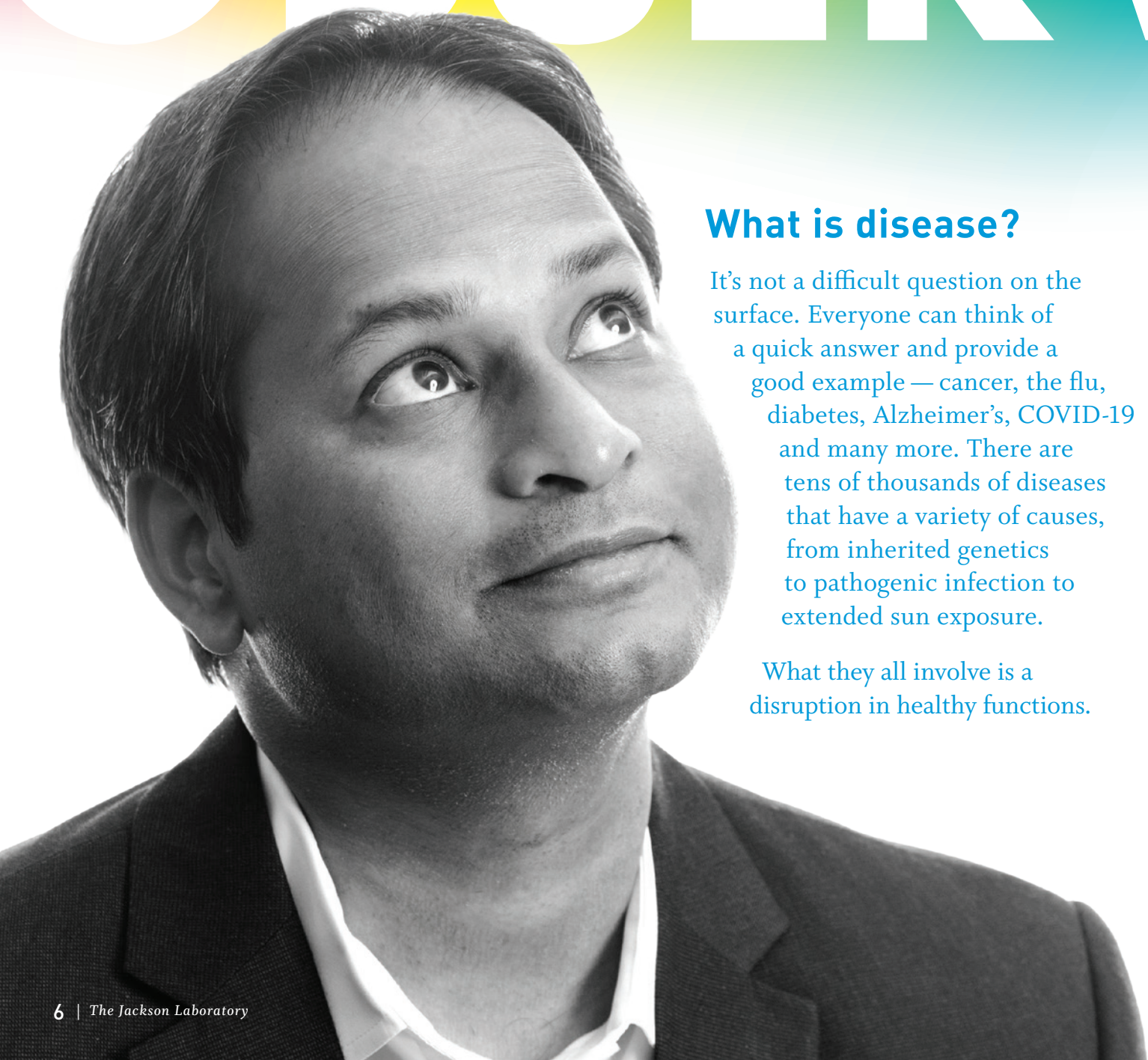
Modern science is simultaneously overwhelming our brains while developing ways to augment their capabilities. It floods us with far more data than we can process, which contains signals and patterns that we cannot detect on our own. Researchers and engineers are therefore creating and implementing computational tools that are able to sift through massive data sets and reveal what they actually contain.

It is like a new form of perception, a way to “see” and learn in ways well beyond what we can do on our own. This form of seeing can be literal, as with data visualization, which represents data in graphical form, such as how a genome loops in three dimensions or a dust cloud extends into space. It can also be indirect, as with machine learning: computerized networks and algorithms “taught” with massive amounts of data to pick out patterns and predict outcomes at a scale and with reliability far surpassing human potential.

Leaders at JAX are playing important roles in the effort to expand our limits and gain new insights and understanding with computational assistance. Their work and achievements are resulting in tools that will have profound benefits in both research and medicine, and, ultimately for patients as well.

# BEYOND OBSERVING

BY MARK WANNER | PHOTOGRAPHY BY MARIE CHAO



## What is disease?

It's not a difficult question on the surface. Everyone can think of a quick answer and provide a good example — cancer, the flu, diabetes, Alzheimer's, COVID-19 and many more. There are tens of thousands of diseases that have a variety of causes, from inherited genetics to pathogenic infection to extended sun exposure.

What they all involve is a disruption in healthy functions.

To really understand disease, however, we first need to thoroughly understand healthy function. That's where the research of JAX Associate Professor Vivek Kumar, Ph.D., comes in.

## Beyond observing

At first, Kumar's work may not seem to connect directly with such broad objectives for human health. He is a behavioral researcher who primarily uses mice to study the genetics underlying addiction. His findings depend on obtaining accurate data regarding mouse susceptibility to addictive behaviors with substances such as cocaine.

Behavioral assays — experimental measurements within specific parameters — provide a snapshot of responses to a situation or stimulus. High-resolution video can track mice movements 24/7, but current assays have

limited accuracy and reproducibility and the video data needs to be watched by experts to actually characterize what the mouse is doing — a labor- and time-intensive process.

“We need to be able to accurately and holistically analyze animal behavior,” says Kumar, “and be able to link that behavior with physiology instead of looking at the two as separate. To accomplish that, our tools need to be more sensitive, higher throughput and require less human involvement throughout the process.”

To achieve his goals, Kumar has turned to machine learning. The trained computational network can analyze thousands of hours of data and tell Kumar whether the animal is carrying out a specific behavior: something that's impossible for humans to do. Furthermore, with altered input, the network can be used to identify different behaviors, providing a general solution for animal behavior detection.

“This kind of ‘behavior extraction’ can provide insight into disease states,” says Kumar. “For example, grooming is associated with anxiety, stress and stereotyped behaviors seen in psychiatric conditions such as autism spectrum disorder and obsessive compulsive disorder. But to determine which genetic variants and molecular pathways are associated with grooming frequency, we need to observe enough mice and acquire enough data to start making functional associations. Before machine learning methods, this was extremely challenging.”

### Grooming, sleep and gait

Kumar’s work with ML has led to a series of papers over the past nine months, beginning in the summer of 2021 with a grooming analysis. That paper went far beyond simply presenting the ML tool, as Kumar took full advantage of the mouse resources at JAX. His lab characterized the grooming behaviors of 62 different mouse strains and thousands of mice. The large data set from different strains displayed a continuum of grooming behaviors and frequencies, with recently wild-derived strains grooming far more than their long-domesticated laboratory counterparts. The researchers were also able to parse out the genetic differences underlying the variability in grooming behaviors. The genes highlighted by the study are known to regulate nervous system function and development and have been implicated in

neurodegenerative diseases. Furthermore, Kumar was able to link the underlying genetics of grooming behavior in the mouse with human psychiatric traits, providing new cross-species insights.

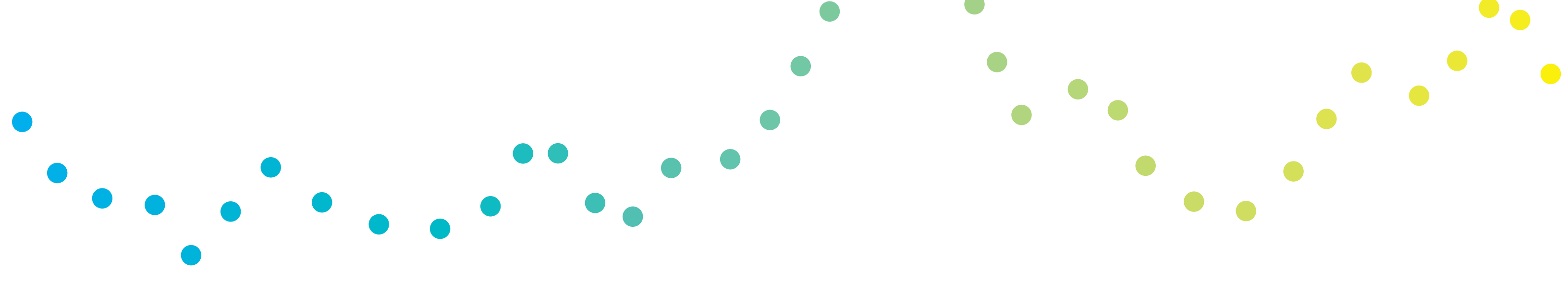
Using other ML approaches, subsequent papers from Kumar’s lab have analyzed mouse movement — gait and pose — and sleep patterns. As with grooming, both are closely tied with human health. Sleep disruption has been associated with a growing list of health consequences, including increased risk for Alzheimer’s disease, and altered movement can be extremely important for insight into multiple diseases.

In one striking example, Kumar analyzed the movement of a Down syndrome mouse model that was created by another JAX faculty member, Muriel Davisson, Ph.D., almost 20 years ago. Children with Down syndrome often have motor changes and can have less coordination.

“Our data shows that Down syndrome mice move very differently than control mice. Our methods are simple and scalable because we watch the animal move around naturally, and these approaches could potentially be used to screen for therapeutic drugs,” says Kumar. His team also found movement changes in mouse models of autism spectrum disorder, Rett syndrome and amyotrophic lateral sclerosis.

To determine which genetic variants and molecular pathways are associated with grooming frequency, we need to observe enough mice and acquire enough data to start making functional associations. Before ML methods, this was extremely challenging.

Vivek Kumar



The sleep analyses also represent an important step forward for the field. It's possible to get accurate sleep data, particularly using electroencephalogram/electromyogram outputs to determine sleep states. But EEG/EMG requires invasive physical implantation of electrodes in the mice as well as expert analysis of EEG/EMG readouts, both of which introduce unwanted variables and severely limit the scalability of the research. Kumar collaborated with Allan Pack, MBChB, Ph.D., at the University of Pennsylvania to create a sleep state classification system that is noninvasive and high throughput. It can accurately score wake, non-rapid eye movement and rapid eye movement sleep states. While only one strain of mouse was used for the study, it can now be applied to genetically diverse mouse strains to map the genetic architecture of sleep for the first time.

### Surpassing human capability

Much of what Kumar has accomplished with ML has enhanced human expertise, capturing more data and insight with equivalent accuracy. With gait and sleep, however, he has moved the tools beyond real-time human capability, and is able to detect differences more subtle than can be seen with the naked eye. His ultimate goal is to use advanced phenotyping methods “the process of measuring biological traits” to understand how altered genetics and neurobiology lead to a wide array of diseases, and to screen for better drugs for human diseases that are modeled in the mouse.

Kumar has expanded his studies by partnering with the JAX Aging Center and Professor Gary Churchill, Ph.D.

“We’re working to visualize ‘frailty’ for aging research, to accurately determine the biological — not calendar — age of an animal,” says Kumar. “Frailty is currently the best method to quantitate biological age, and even an experienced mouse technician is poor at assessing biological age, so it’s more abstract and more difficult than what we’ve tackled so far.”

What’s the payoff? The methods remove human bias and provide a scalable platform for interventional studies of aging. “Imagine being able to use the mouse to screen for genetic or pharmacological factors that enhance healthy aging as opposed to unhealthy aging,” says Kumar. “Or being able to detect signs of aging earlier than current methods for better preclinical animal studies.” The potential benefits are enormous.

### Toward human health

Of course, Kumar is working with mice, and the ultimate goal for any kind of disease research is to translate experimental findings to clinical progress. Mice have proven to be invaluable for learning about fundamental mammalian biology, but bridging the gap to human medicine, with our highly variable genetics, environments and behaviors, is far more difficult. In order to improve the mouse-human interface and also improve medical safety and efficacy, it’s essential to obtain better preclinical data. That’s the whole point of Kumar’s research, and he returns to autism spectrum disorder as an example.

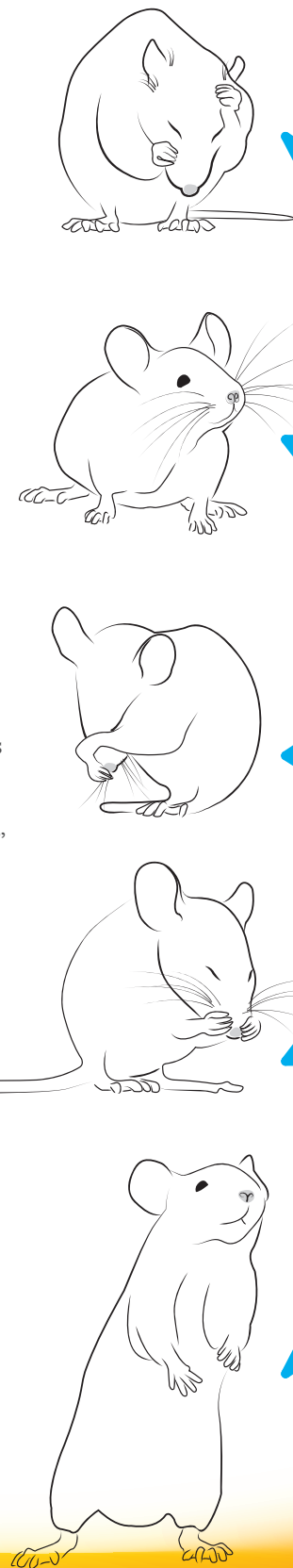
“In autism, the earliest symptoms are motor symptoms, and the cognitive changes develop later,” he says. “Cognitive functions related to ASD are very hard to measure in mice, so what if we use the motor symptoms for preclinical research instead? If you’re screening compounds for early ASD intervention, looking at mouse movement with the ML tool could be a great surrogate.”

# INPUT

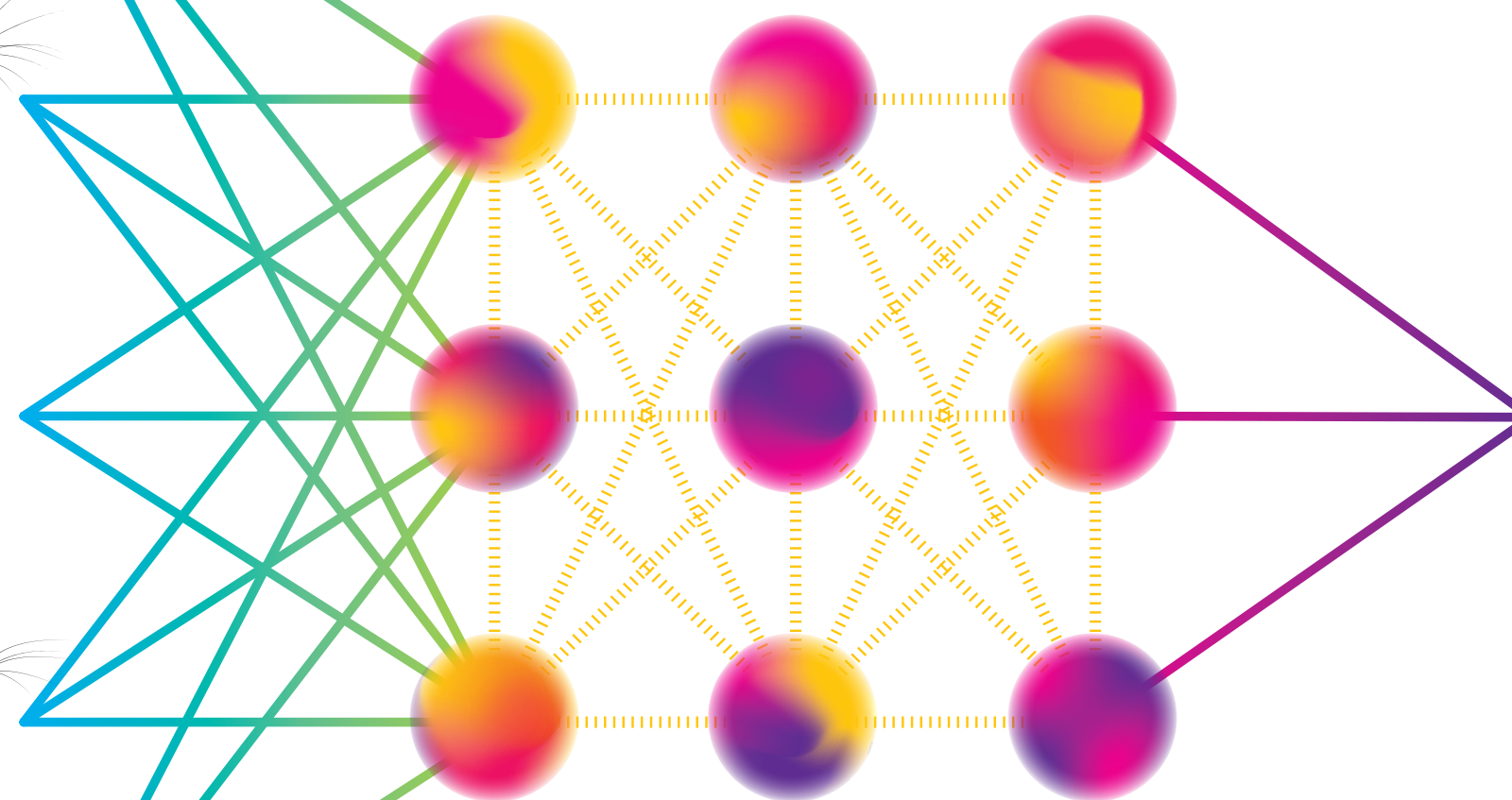
Kumar originally conceived of the tools in the context of behavior and addiction, but they can apply to diseases of all kinds beyond those already mentioned, including complex diseases that need better treatments. Identifying and quantifying the small changes associated with diabetes, cancer, Alzheimer's disease and more will narrow the focus on the genetics and molecular processes involved. A solid foundation of preclinical data obtained in concert with human patient data will provide a much better starting point for identifying effective therapeutic targets as well as weeding out those that are unlikely to work. The ability to identify changes in the mice before overt symptoms emerge also holds promise for improving diagnostics and opening avenues for early intervention.

The broader concept is still relatively new. "Some 'preaching' will need to happen," says Kumar. "But automated ML-based mouse phenotyping is the future for the field. In 10 years every cage will have video capture."

It will be exciting to see what the next 10 years bring.

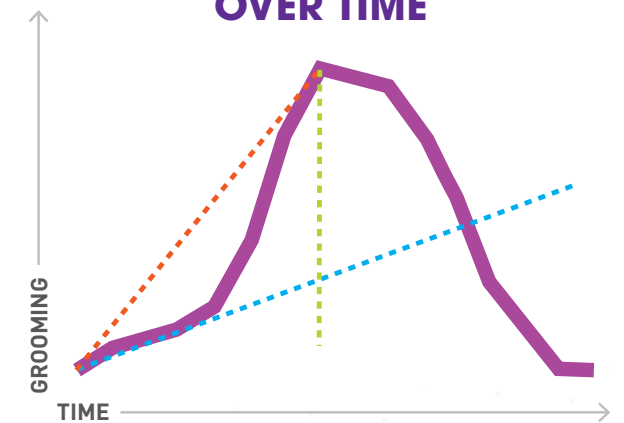


# PROCESSING



# OUTPUT

## ANALYSIS OF GROOMING BEHAVIOR OVER TIME



In machine learning, one part of the larger field of artificial intelligence, data is used to "teach" a computational algorithm to perform a specific task. The data (input) in this case are video frames showing mice grooming or not grooming, as identified by expert observers. It is then processed in a neural network (processing), and the algorithm is refined. Once trained and tweaked, the network can process large amounts of data automatically and provide accurate analyses (output).

# 5 QUESTIONS WITH MATTHEW GERRING

BY SOPHIA ANDERSON  
PHOTOGRAPHY BY TIFFANY LAUFER

**Q: What defines machine learning?**

**A:** Machine learning is based on statistics. It generates models from training data and applies those to novel data sets, resulting in predictions.

**Q: Why is there so much excitement around machine learning today?**

**A:** People have now noticed machine learning in the wider community. They've noticed how good it is because they have it on their smartphones, and they're able to identify butterflies, birds and plants, things like that. But these models have existed in mathematics for a long time and they've just gotten better.

**Q: How is JAX using machine learning in research?**

**A:** The Kumar laboratory, for one, determines mouse behavior via video, and deep learning and machine learning can go further by analyzing those videos and drawing conclusions about individual mice.

**Q: What is the process of establishing a machine learning model?**

**A:** Researchers make interpretations, and each of those experts provides data for input into your model. You'll get a virtual version, but each will have their strengths and their biases. Machine learning models need constant refining.

**Q: What do you envision for the future?**

**A:** You can imagine what it would be like by watching science fiction films where they are able to talk to the computer, presumably using something like natural language processing. They get the computer's opinion based on the different models that the computer has about reality. The individual software components on which we are now working will add up to that capability one day.



Computational sciences Senior Manager Matthew Gerring



# COMPLEX PROBLEMS, COMPUTATIONAL SOLUTIONS

BY MARK WANNER

Faced with difficult challenges in their research, JAX principal investigators develop their own computational tools to overcome them.

As the genomics era moves forward with ever greater sequencing capacity and more and more variety in the types of genomics studies performed, we're facing a problem.

The vast amounts of data already produced and still being generated could yield even more insights if they are brought together, but integrating them is difficult because different experiments using different methods have very complex relationships. How can you combine those data and advantage of the diverse results and find the convergent signals in the noise?

My answer is computational tools and methods that allow researchers to integrate data sets from different experiments using different methods and even working with different species.

I developed GeneWeaver, a web-based software system that combines genomics data repositories with powerful analysis tools, to put biologists in the driver's seat and lets them work with big genomics data sets in the ways most valuable to them. Researchers are able to ask questions and compare or contrast disease related signals to find consensus across many different experiments in different populations and species. GeneWeaver currently contains more than 220,000 data sets capturing the role of nearly 2.9 million genomic features in disease associated biology, and its size and power continue to grow as we integrate GeneWeaver with new bioinformatics tools and data services.

**ELISSA CHESLER, Ph.D.,  
professor**

**JULIA OH, Ph.D.,  
associate professor**

Research has shown that our well-being is strongly linked with our microbiome, the multitude of microbial organisms that live in and on us. I investigate how it all works, with a particular focus on the skin microbiome.

Determining how the microbiome affects our health remains an enormous challenge: we must figure out which microbes are present (many remain unknown to this day) and in what numbers, something that is different from person to person and from site to site in or on the same person. And that's just the start.

How does the microbiome interact with host systems? How does it change over time, in different environments, or in disease? How do microbes compete and interact with each other? It all adds up to a lot of data and a lot of analyses.

My group develops experimental and computational tools to make deeper dives into the microbiome and to understand its mechanisms. For example, these tools help researchers identify microbes using sequencing data, determine how quickly they grow, predict their function in the body, figure out what impact a microbe might have on the skin, and more. Our work provides insights that we need to better understand how our microbiomes work with us and, when disrupted, how they can contribute to our own dysfunction and disease.

# PUTTING MACHINE LEARNING TO WORK FOR CANCER PATIENTS

A Jackson Laboratory research team is boosting the speed and accuracy of tumor sample image analysis, for multiple cancer types, offering the potential for more targeted treatment direction without additional testing.

BY JOYCE DALL'ACQUA PETERSON | PHOTOGRAPHY BY DOMINICK REUTER & MARIE CHAO

A patient with cancer is unlikely to meet the pathologist who has analyzed her biopsy sample, yet the pathologist's analysis will drive many decisions around her treatment options.

Pathologist Todd Sheridan, M.D., describes his job this way: "I look through a microscope to evaluate slides containing tissue from a biopsy or resection. Usually this is a histopathology image that we call H&E because the sample is treated with two histological stains, hematoxylin and eosin." Pathologists are trained to recognize cancer cells, he says, "and we have a number of techniques to improve our diagnostic accuracy such as immunochemistry, which uses antibodies to confirm the presence of certain markers on the surface of cancer cells."

## Improving tumor imaging and interpretation

But, Sheridan says, pathology has lagged behind radiology and some other medical fields in deploying advanced computer image analysis to improve tumor sample interpretation. Sheridan holds joint appointments with Hartford Pathology Associates at Hartford Hospital in Connecticut and The Jackson Laboratory for Genomic Medicine in Farmington, where he works with Professor Jeffrey Chuang, Ph.D., on a project to harness the power of machine learning for pathology.

Chuang was the lead author of a December 2020 study published in *Nature Communications* that harnessed the power of machine learning to capture detailed information about cancers from scanned images. From The Cancer Genome Atlas, they used 27,815 scanned images from over 19 cancer types to "train" their convolutional neural network software to reveal how cancers from different organs are related.

Today researchers around the world are beginning to explore the possibilities of computer image analysis, including Chuang's collaborators at Yale University, Boston University and University of Massachusetts Boston. His lab at JAX, he says, brings special capabilities to the challenge.

"One of our leading strengths is our ability to integrate data from multiple cancer types simultaneously," Chuang says. "For example, we have found image features that can predict cell density in tumors, and these features are accurate in breast tumors, lung tumors, bladder tumors and many others. In addition, we have special expertise in combining protein imaging data with histopathological images, which allows us to make better clinical predictions, and also identify those cases where the H&E data are already capturing the key predictive signals."

If we can provide pathologists with tools to interpret tumor images faster and more precisely, cancer patients will benefit from more targeted and effective treatment approaches.

Jeffrey Chuang



Cancer is a severe disease that requires accurate diagnosis. Histology slides, thin slices of stained tissue examined under a microscope, are an essential tool in cancer diagnosis. However, cancer cells can look very similar to healthy cells, making it difficult to identify cancerous tissue quickly. Computer-aided diagnosis is becoming increasingly common in cancer care, as AI can often achieve better accuracy than human experts in a fraction of the time.

Hematoxylin stains cell nuclei a purplish blue, and eosin paints the extracellular matrix and cytoplasm pink with other structures taking on a mix of the two shades. This allows scientists — and now AI programs — to differentiate between parts of a cell easily.

### Leveraging the power of machine learning against cancer

Now the team, including machine-learning expert Ali Foroughi Pour, Ph.D., is zeroing in on identifying a common hallmark of breast, gastric and other cancers — directly from H&E images.

Pathologists can tell a lot from an H&E image, Chuang says, “such as whether cells are dying or whether a tumor has an unusual appearance. But they can’t ‘see’ the genetic aberrations that are driving the cancer, such as the amplification of the HER2 gene that is common in breast cancer, unless they use a special HER2 stain as well as the H&E.”

When a pathologist identifies this HER2 amplification, he says, the patient is usually treated with targeted agents such as the drug trastuzumab (better known as Herceptin). “This treatment can be very effective for these patients, and lead to very good outcomes,” Chuang says.

Working with collaborators at the Yale University School of Medicine, the Chuang lab analyzed about 200 H&E samples for HER2 amplifications. After correcting for some variations in image processing, he says they achieved an accuracy of about 80%. “What’s remarkable about this is that pathologists cannot identify HER2 status directly from H&E images. Now, we can,” he says.

Jill Rubinstein, M.D., Ph.D., has joint appointments in surgical oncology at Hartford Hospital and computational oncology in the Chuang lab. “What Todd and I bring to the team is our clinical perspective. As we build tools from these data sets, how are we going to shape those insights that we get from deep learning into a clinical tool that might actually impact patient care?”

### A new wave of advanced data types

Today, Rubinstein points out, genomic sequencing of tumors accurately identifies genetic mutations that drive cancers. “High-throughput sequencing provides a very detailed look into what tumors are doing, how they are evolving and what pathways are driving them. But it’s expensive to sequence tumors, and not everyone has access to that kind of care.”

Instead, she says, “if we can marry molecular data and deep learning-based imaging data, those H&E-stained images that are already produced during a standard clinical workflow could provide valuable information for clinicians. These tools can be implemented without any additional assays being performed on the tumor. You can implement them remotely, so the patient doesn’t have to be at a center that has specialized capacity for molecular testing.

“Bringing high-level computational insights to any patient, without additional cost, would be a great leveler,” Rubinstein says.

Chuang says the team is already bringing this data-driven image analysis approach to other tumor types, including colon cancer, melanoma and sarcomas.

“The thing about this kind of imaging data,” he says, “is that there’s a natural intuition to interpreting it. Imaging data tells you the shapes of cells, where they are, what kinds of cells are active, whether there are dead cells or big spaces between them. But there may be very important mechanistic things in the cells that we just can’t see.” Sequence data, in contrast, provides great detail, but “it’s less intuitive.”

Chuang says the machine-learning image analysis techniques his lab is refining are part of a new wave of advanced data types, such as spatial transcriptomics, and spatial protein and metabolomic profiling. “And we’re working on methods to integrate these so medical images can communicate much more to clinicians.”

# glue: MAKING SENSE OF THE UNIVERSE IN 3D

BY MARK WANNER

## RIGHT NOW

**1.5 million kilometers away** the James Webb Space Telescope is delving deep into the universe — as much as **13.6 billion light-years away**, soon after the Big Bang — and revealing far more than what was currently known.

## RIGHT NOW



What connects the farthest reaches of the universe and our own internal universes? In the case of The Jackson Laboratory, it is the aptly named glue solutions inc., a firm specializing in data science and visualization. glue solutions, led by Harvard University Astronomy Professor Alyssa Goodman, Ph.D., grew out of a project to visualize astronomy data, including facilitating data analysis and visualization for the James Webb Space Telescope. Now it is working with JAX to do the same for genomics, providing ways to bring the data to life through visual representations.

### What is data visualization?

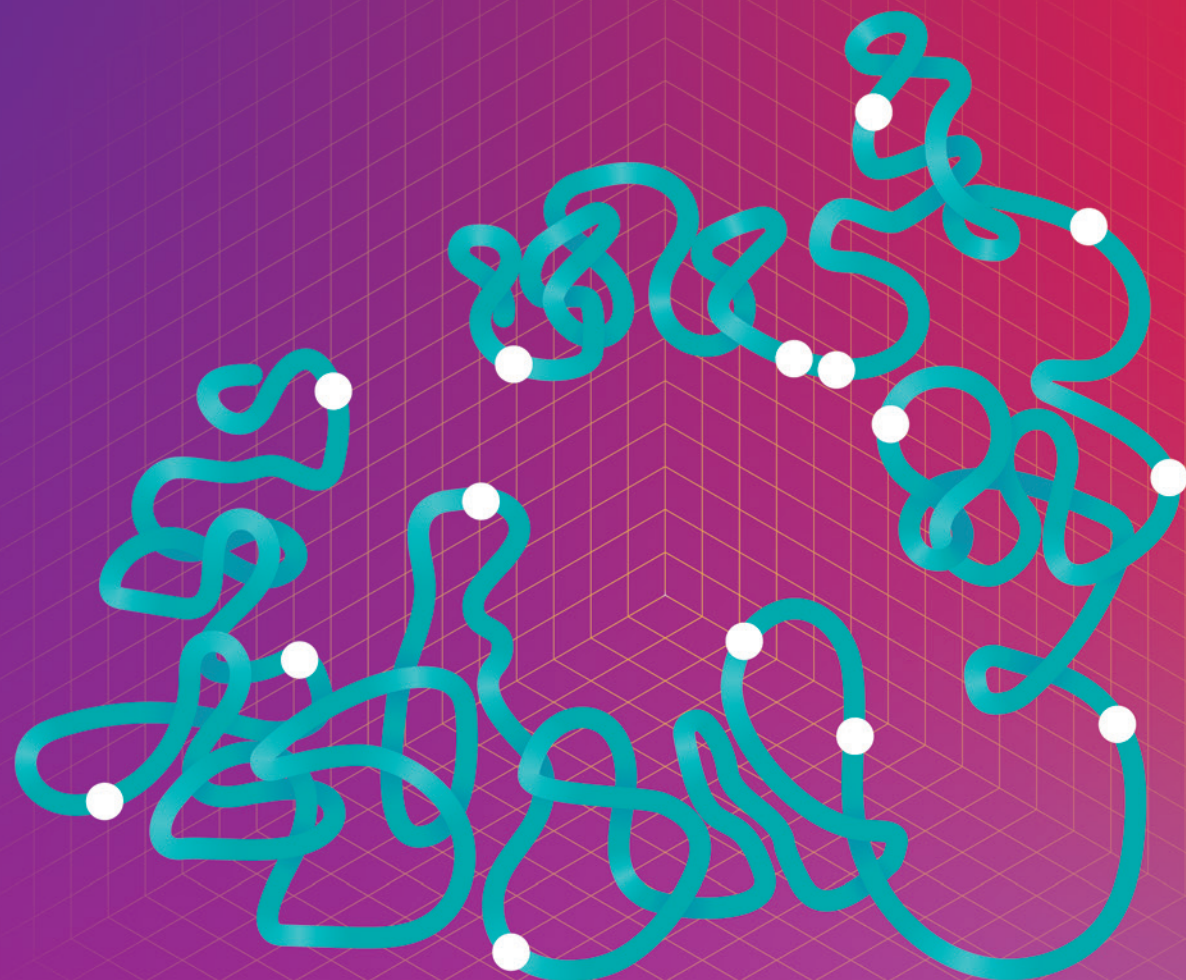
Examples of basic data visualization are all around us. A topographic map that shows elevation lines that make it easy to pick out mountains and follow river valleys is one. The kind of infographic that *USA Today* made famous is another, allowing readers to quickly see what the numbers in an article actually represent. And anyone clicking on the graphing tool in an Microsoft Excel spreadsheet is working with data visualization on their own computer.

So-called big data, with data sets containing millions of data points or more, also benefits from data visualization to better comprehend its signals and communicate them to others. But handling

this amount of data — imagine 32 billion rows (just 10 genomes) in an Excel file! — demands expertise, robust tools and a lot of computational horsepower. And without visualization, the data can be difficult or impossible with which to work, even for experts. Astronomy provides an excellent example, as its data is a vast array of signals from different directions in different wavelengths and energy levels, all of which must be assembled, analyzed and represented visually in order to “see” the features of the universe around us.

### From molecular clouds to molecular bases

glue has already had a significant impact on astronomy data visualization. Using data from Gaia, the Global Astrometric Interferometer for Astrophysics, a European Space Agency astronomical observatory mission, astronomers have been studying two molecular clouds, Perseus and Taurus. In astronomical terms, the clouds are relatively near to Earth — as close as 400 (Taurus) and as far as 1,300 (Perseus) light-years away — and have been well studied in two dimensions. Using glue-powered visualization and augmented reality, however, researchers mapped them in three dimensions for the first time last year. The visualizations have provided valuable new insight into how the clouds formed (likely from the same



## RIGHT NOW

# RIGHT NOW

**0 kilometers away,** researchers are delving into our genomes — all 3.2 billion base pairs — and extending their analyses far beyond the linear sequences of our genomes to reveal what underlies health and disease.

supernova shockwave), their relative densities and the dynamics of star formation within the clouds.

The story of Perseus and Taurus, or PerTau for short, is the sort of thing that the glue software package was initially developed to do. That said, it was also designed from the ground up to be an open source, multi-disciplinary tool. When launching the project in the late 2000s, Goodman and colleagues sought to make it applicable to a wide variety of data types.

“We developed the predecessor to what is now glue based on medical imaging software used for surgery planning in three dimensions,” says Goodman. “Then, back in 2012, we were offered unexpected James Webb Space Telescope funding based on a paper, only half-written at the time, designed to explain what glue ‘could’ do — even before the software was really ready for widespread use.”

The glue project has grown rapidly since, and its open source code is being used for many different purposes. In fact, the situation became unwieldy — “fun projects, but too many requests,” Goodman notes — so the team launched glue solutions, inc. to work with commercial partners. Though for-profit, glue solutions gives away half of its proceeds to support the open source software community.

glue is now being applied to genomics at JAX through a suite of tools called “glue genes,” making visualizations possible for the endless streams of bases and biological complexity to reveal patterns and insight otherwise hidden in the data.

“In astronomy, our high-resolution images often come with a third dimension, creating so-called ‘image cubes,’” says Goodman. “In biology, high-resolution 3D data, preserving information about where gene sequences come

from spatially within a sample, is about to be available. Looking toward the future, in a discussion a couple of years ago, the former JAX President and CEO [Ed Liu, M.D.,] and I talked about how high-resolution ‘spatial transcriptomics’ would soon be possible, and that original conversation has led to glue genes. Soon, the augmented reality functionality glue offers astronomers will also come to glue genes.”

## Genomics revealed

Data visualization for genomics data is nothing new. Genome browsers, which make analyzing linear sequences much easier, are important tools for annotating genome sequences. For example, locating known protein binding sites, or epigenetic marks (chemical compounds that change function without changing the sequence) on the sequence. Another common tool is a heat map, which uses different colors and color intensities to represent relative levels or magnitudes in the data and biology. They can be used to represent the levels of gene expression for different genes across a variety of samples. There are also circos plots, the often highly complex circular diagrams that can show various types of genomic data at specific locations, as well as the interactions between them.

But genomics is moving beyond linear sequences. The chromosomes themselves function in three dimensions, and sequences that may be far apart on a linear basis can frequently interact, directly or indirectly, in ways that are important for gene expression, regulation and function. Visualizing genomes in three dimensions provides the opportunity for new insights that linear sequences simply don’t provide. glue and JAX are working to leverage the capabilities of glue in an entirely new way.

# A gift for the next generation

“One of the great limitations we see in the biological sciences is the ability to link, integrate, explore and visualize different types of highly complex genomics data generated from different technology platforms,” says Chief of Presidential Initiatives Madeleine Braun, Ph.D., who is coordinating the collaboration at JAX. “Together with Dr. Alyssa Goodman and her team, we’re adapting glue solutions’ tools to address this urgent need, making it easier to bridge, stack up and cross-analyze diverse types of data to find new insights.”

## Visualizing faint signals

JAX’s work with glue comes at a time when JAX researchers are facing some of the most difficult challenges in biological research head on. An important example is a study into Type 2 diabetes, a disease that involves a collision of genetic, environmental and behavioral factors. A team at JAX is working to merge the signals found in human studies, which have identified hundreds of genomic locations associated with T2D, with research using genetically diverse mice.

“The Type 2 diabetes study has been a significant scientific and operational effort for the Laboratory,” says JAX Principal Scientific Software Engineer Jake Emerson, “We’re working to develop an effective interface between human and mouse data across the

research program, and to make the data readily accessible for all researchers at JAX. glue genes is an exciting part of that effort, providing intuitive data visualization up front as well as the ability to extend the toolkit for specialized inquiries moving forward.”

Emerson and his team are working to integrate a collection of biological data types to reveal the full picture. It involves the genetic code, the messages transcribed (which can be processed in different ways), the proteins produced and more, covering the full arc of function. Ultimately, the molecular data can be merged with the measurable physical and physiological traits that result and show how, in disease, function is disrupted.

So, as the space telescopes help scientists understand the universe around us, glue technology will play an equally important role in helping researchers understand our own internal universes with precision and high resolution. The first step is to apply the lessons learned in the T2D pilot to other research areas, such as cancer, Alzheimer’s disease, autoimmune disease, and other difficult-to-treat diseases and conditions. The ultimate goal? It may seem distant now, but a future with better health and less disease may be closer than you think.

Inspired to encourage budding scientists toward

**The Jackson Laboratory’s Summer Student Program,**

**Lynn Moorhead Riddiford, Ph.D., ’53, ’54,** a renowned

entomologist and biologist, hopes her gift in support of a JAX Living-Learning Community for rising high school seniors will unlock the hidden potential in a new generation of students.

When fully funded, the JAX Living-Learning Community will provide students under the age of 18 with dedicated summer housing and special programming throughout their final year of high school, including guidance on the college admission process, scholarship applications and career coaching.

Lynn and her husband, Jim Truman, Ph.D., hope their support of the summer experience in Bar Harbor — given to JAX through a bequest intention — will move others to follow their example. Read the full story and learn how to join the effort at [www.jax.org/give](http://www.jax.org/give).

**It’s easy to name JAX as a beneficiary of your retirement plan, life insurance, policy, will or trust. Learn more at [www.jax.org/giftplanning](http://www.jax.org/giftplanning).**

**We are happy to answer any questions you may have.**

**Please contact us at [planmygift@jax.org](mailto:planmygift@jax.org) or call (860) 837-2328 or (800) 474-9880.**



2019 SUMMER STUDENT COMMENCEMENT AND SYMPOSIUM | PHOTOGRAPHY BY TIFFANY LAUFER



600 Main Street  
Bar Harbor, ME 04609-1523

Forwarding service requested

A PUBLICATION OF THE JACKSON LABORATORY  
[www.jax.org](http://www.jax.org)

