An enterprising Whitehead Fellow exposes new principles of biology. The discoveries have boosted our understanding of health and disease, and led to the development of many drugs and therapies. So these individuals really should be household names.

In the household that is Whitehead Institute, however, several of the recipients’ names are well known, indeed. There’s Rockefeller University professor and Howard Hughes Medical Institute investigator Cornelia Bargmann, who, as a graduate student in the 1980s, trained in Weinberg’s lab. And, there’s Eric Lander, a former Whitehead Fellow and Whitehead Member who went on to become Director of the Broad Institute. Finally, we have renowned geneticist David Botstein, now at Princeton. Years ago, as a professor of genetics at MIT, Botstein mentored two promising young scientists: Lander and current Whitehead Institute Director David Page.

It’s impossible to predict whether Whitehead will have such strong connections to future Breakthrough Prize winners, but with links to more than one-third of the first-timers, we’re content—for now—to savor this historic moment.

WINNING COMBINATION: Whitehead Founding Member Robert Weinberg, right, and former Whitehead Member Eric Lander toast each other after each winning a new Breakthrough Prize in Life Sciences.

Photo: Ceal Capistrano, Whitehead Institute
**RESEARCH NEWS**

**NUTRIENT-SENSING ENZYMES KEY TO STARVATION RESPONSE AND SURVIVAL IN NEWBORN MAMMALS**

In the perinatal hours immediately after birth, a newborn mammal faces the sudden loss of food supply from its mother. Under normal circumstances, newborns mount a metabolic response to ward off starvation until feeding occurs. This survival response involves a process of controlled breakdown of internal energetic sources known as autophagy. Although autophagy has been well documented, the key mechanistic regulators of autophagy in vivo have remained poorly understood.

Whitehead Institute researchers have discovered that a family of nutrient-sensing enzymes, dubbed Rag GTPases, modulates the activity of the mTORC1 protein complex, whose inhibition is essential for autophagy and survival in newborns. The finding, reported late last year in the journal Nature, emerges from the lab of Whitehead Member David Sabatini, whose earlier research on mTORC1 has been well documented, the key mechanistic regulators of autophagy in vivo have remained poorly understood.

Whitehead Institute researchers have discovered that a family of nutrient-sensing enzymes, dubbed Rag GTPases, modulates the activity of the mTORC1 protein complex, whose inhibition is essential for autophagy and survival in newborns. The finding, reported late last year in the journal Nature, emerges from the lab of Whitehead Member David Sabatini, whose earlier work on mTORC1 has been well documented, the key mechanistic regulators of autophagy in vivo have remained poorly understood.

To assess the impact of this Rag GTPase-mTORC1 relationship in mammals, the lab generated mice genetically altered to continuously express an active form of the GTPase RagA and compared them with wild-type mice. In normal mice, RagA is activated in the presence of nutrients, and turns on the mTORC1 pathway, which regulates organismal growth in response to nutrient availability. If the mice are starved, the mTORC1 activity of the RagA enzyme always on is pretty shocking," says Sabatini, who is also a professor of biology at MIT and a Howard Hughes Medical Institute (HHMI) investigator. "A normal neonate animal within an hour after birth responds to that condition, but one with its RagA stuck ‘on’ doesn’t, and it dies. It basically has a huge energetic and nutritional crisis because it can’t make the adaptations.”

"These striking results stunned Alejo Eleyan, a postdoctoral researcher in the Sabatini lab, and first author of the Nature paper that describes this work.

“We were surprised that there was no inhibition of this pathway independent of RagA—there is no backup system, says Eleyan. "And that RagA is a more global nutrient sensor than goes beyond its known function as an amino acid sensor." RagA’s role as an amino acid sensor had been established in cultured cells by the Sabatini lab. Yet when Eleyan compared nutrient levels in fasting newborn RagA-active mice with those of fasting pups with normal RagA, not only amino acids were reduced in RagA-active animals, also glucose levels were dangerously low. The animals were unable to "sense" either of these reductions, so autophagy failed to initiate in the RagA-active pups, which all died within hours of birth.

This newly identified function for RagA suggests much remains unknown about the cell biology of nutrient sensing, an area of research that Sabatini and his lab continue to investigate.

**STUDY REVEALS RATE AT WHICH KEY GENETIC DELETIONS CONTRIBUTE TO MALE INFERTILITY**

A large-scale analysis of Y chromosomes from more than 23,000 men finds that two spontaneously recurring deletions along a complex region of the Y chromosome are responsible for approximately 8% of cases of failed sperm production. Although previous research had identified deletions in the region of the Y known as AZF (for azoospermia factor c) as causing severe spermatogenic failure (SOF), this latest analysis, conducted by Whitehead Institute Director David Page and colleagues, is the first to determine how prevalent these deletions are in the general population.

According to the study, published recently in the American Journal of Human Genetics, the deletion known as ba/bα is found in one of every 4,500 men, increases the risk of SOF 144 times, and is responsible for roughly 6% of cases.

"This deletion almost always results in spermatogenic failure, so it would be extremely rare for it to be transmitted from father to son without medically assisted reproduction," says Page. "Because of this, we can conclude that its prevalence in the population essentially reflects the rate at which this deletion arises spontaneously in men.”

"Medically relevant population genetics studies are well established for most of the human genome, and these areas are likely to be as unstable and prone to mutation as those on the Y chromosome. While the effects of the known deletions of the AZF region appear to be limited to sperm production, substantially more harmful health effects are apt to arise from mutations elsewhere. Given the inherent challenges of obtaining accurate and complete DNA sequences of mirrored regions, Page believes that the current reference sequence of the human genome is missing potentially meaningful detail—and that the time has come to apply SHIMS broadly.

"The key to SHIMS starts with the realization that there are areas of the human genome that are almost perfectly mirrored repeated sequences that are greater than 99% identical," says Page, who is also an investigator of the Howard Hughes Medical Institute. "When you assemble a sequence from multiple unrelated chromosomes, as was done with the human genome, you cannot make sense of minute but substantial differences.”

"The human genome reference is a consensus sequence, which is a politically wonderful outcome," he adds. “But in mirrored regions,
The main role of mature Sertoli cells is to provide support and nutrition to the developing sperm cells. Furthermore, Sertoli cells have been demonstrated to possess trophic properties, which have been utilized for the stimulation of non-testicular cellular grafts in transplantsations. However, mature Sertoli cells are histiotically inactive, and the primary immature Sertoli cells during prolonged culturing became degenerate in the Petri dish. Therefore, finding an alternative source of these cells is of independent interest both for basic research and clinical applications. The idea is if you could make Sertoli cells from a skin cell, they will become suitable for supporting the spermatogenesis process when conducting in vitro fertilization assays or protecting other cell types such as neurons when co-transplanted.

Next the scientists devised a cocktail of five transcription factors that activate the epithelial cells’ embryonic Sertoli cell genetic program. The resulting cells exhibited many of the characteristics of embryonic Sertoli cells, including aggregating, forming tubular structures similar to the seminiferous tubules found in the testis, and secreting the typical Sertoli cell factors. When injected into a mouse fetal testis, the trans-differentiated cells migrated to the proper place and integrated into the endogenous tubules. Overall, the injected cells behaved like endogenous embryonic Sertoli cells, despite expressing a few genes differently. “The trans-differentiated cells were closely interacting with the native germ cells, which shows that they do not have any bad effect on the germ cells,” says Yossi Buganim, a postdoctoral researcher in the Jaenisch lab and first author of the Cell Stem Cell paper. “Instead, they enable those germ cells to survive.”

In fact, when the embryonic Sertoli-like cells were allowed to sustain other cells in a Petri dish, Buganim noted that the cells supported by the trans-differentiated cells thrived, living longer than cells sustained by actual native Sertoli cells. Encouraged by these results in vitro, Buganim says he would like to investigate whether the embryonic Sertoli-like cells retain this enhanced supportive capacity after transplantation into the brain, where the cells could sustain oligodendrocytes. If so, they could have applications in the development of neuron-based therapies for neurodegenerative disorders such as ALS and Parkinson’s disease.

In another finding, the researchers used the genome-editing tool CRISPR to manipulate the genomes of skin cells into embryonic Sertoli-like cells. They were then able to transplant the cells back into the testes of mouse models. “This is one of the rare cases where we can explain a normal human-to-human variation,” says Lodish, who is also a professor of biology and director of MIT’s Broad Institute. “I think it’s a window on human evolution. Why this should have happened, we have no idea, but it does.”

Lodish likens cyclin D3’s role in RBCs to that of a clock. In some people, the clock triggers RBC progenitors to mature after four rounds of cell division, resulting in fewer but larger RBCs. In others it goes off after five cell division cycles, which leads to production of a greater number of smaller RBCs. In both cases, the blood usually has the same ability to carry oxygen to distant tissues.

The initial hint of cyclin D3’s importance came from GWAS, genetic surveys of large numbers of people with or without a particular disease. Researchers compare the groups in an attempt to identify genetic variations.

“The problem with most GWAS is that you get a bunch of potentially interesting genes, but that doesn’t tell you anything about the functional biology, so you really have to figure it out,” says Leif Ludwig, a Lodish graduate student and co-author of the Genes and Development paper. “You only know something has a role, but you don’t know how it can cause variation. This work on cyclin D3 is a really nice example of how functional follow-up, especially on a gene like GWAS association can really teach us something about underlying biology.”

In the case of RBC size and number, a mutation affecting cyclin D3 production bubbled to the surface from the GWAS’s murky genetic data. Ludwig and co-author Vijay Shankaran then confirmed that reduced or inhibited cyclin D3 expression in mice and in human RBC progenitors caused these cells to halt cell division and mature earlier, producing larger and fewer red blood cells than mice and cells with uninhibited cyclin D3 production.

As one of only a handful of studies that have successfully used GWAS to produce definitive biomedical results, Shankaran is excited that this work confirms the value of such genetic studies.

“Can genetics teach us about biology?” asks Shankaran, also a postdoctoral researcher in the Lodish lab. “Yeah! This work tells us that as genetic studies identify new genes, there will probably have been a lot of things biologists have ignored. Genetics allows you to shine a spotlight on something interesting and then home in on it see what can be learned.”

Researchers at Whitehead Institute and Memorial Sloan-Kettering Cancer Center have defined and analyzed the crystal structure of a yeast Argonaute protein bound to RNA. This complex plays a key role in the RNA interference (RNAi) pathway that silences gene expression. Describing the molecular structure of a eukaryotic Argonaute protein has been a long-standing goal of the RNAi field for close to a decade.

Lodish likens cyclin D3’s role in RBCs to that of a clock. In some people, the clock triggers RBC progenitors to mature after four rounds of cell division, resulting in fewer but larger RBCs. In others it goes off after five cell division cycles, which leads to production of a greater number of smaller RBCs. In both cases, the blood usually has the same ability to carry oxygen to distant tissues.
RNAi depends on two proteins, Dicer and Argonaute. Dicer recognizes double-stranded RNA (dsRNA), latches onto it, and chops it into 21-23 nucleotides-long Argonaute recognizes the dsRNA bits, discards one strand, and uses the other as a guide. When a single-stranded RNA matches the guide RNA’s sequence, Argonaute cleaves the target RNA, thereby preventing it from serving as a template for protein production.

To determine the structure of Argonaute, Bartel and graduate student David Weinberg partnered with Kotaro Nakashima in Dinesh Patel’s lab at Sloan-Kettering. Although the team expected to solve the structure of Argonaute alone, they were surprised to find that the protein came along with small bits of RNA that were also observed in the structure. The incorporation of these RNAs had switched the protein into an activated state that contained a four-component active site, the identification of which solved a longstanding mystery of what constituted the “missing” fourth component. With the structure of this complex in hand, scientists now have a better understanding of how it works.

“Seeing the crystal structure of a eukaryotic Argonaute for the first time was very exciting—it’s such a large protein with a complicated topology and many moving parts,” says Weinberg. “It’s a really impressive molecular machine.”

FISHING FOR ANSWERS TO AUTISM

Fish cannot display symptoms of autism, schizophrenia, or other human brain disorders. However, a team of Whitehead Institute and MIT scientists has shown that zebrafish can be a useful tool for studying the genes that contribute to such disorders.

led by Whitehead Member Hazel Sive, the researches set out to explore a group of about two dozen genes known to be either missing or duplicated in about one of every thousand autistic patients. Most of the genes’ functions were unknown, but a new study by Sive and Whitehead postdoc Alicia Blaker-Loe, Sunny Gupta, and Jasmine McCammon, revealed that nearly all of them produced brain abnormalities when deleted in zebrafish embryos.

The findings, published recently in the journal Developmental Cell, should help researchers pinpoint genes for further study in mammals, says Sive, who is also professor of biology and associate dean of MIT’s School of Science. Autism is thought to arise from a variety of genetic defects; this research is part of a broad effort to identify culprit genes and develop treatments that target them.

“That’s really the goal—to go from an animal that shares molecular pathways, but doesn’t get autistic behaviors, into humans who have the same pathways and do show those behaviors,” Sive says.

Sive recalls that some of her colleagues chuckled when she first proposed studying human brain disorders in fish, but it is actually a logical starting point, she says. Brain disorders are difficult to study because most of the symptoms are behavioral, and the biological mechanisms behind those behaviors are not well understood, she says.

“We thought that since we really know so little, that a good place to start would be with the genes that confer risk in humans to various mental health disorders, and to study these various genes in a system where they can readily be studied,” she says.

Those genes tend to be the same across species—conserved throughout evolution, from fish to mice to humans—though they may control somewhat different outcomes in each species.

In the latest study, Sive and her colleagues fo- cused on a genetic region known as zp1261, first identified by Mark Daly, a former Whitehead Fellow who discovered a type of genetic defect known as a copy number vari- ant. A typical genome includes two copies of every gene, one from each parent; copy numb- er variants occur when one of those copies is deleted or duplicated, and this can be associ- ated with pathology.

The central “core” of zp1261 includes 25 genes. Both deletions and duplications in this region have been associated with autism, but it is unclear which of the genes might actually produce symptoms of the disease.

“At the time, there was an inkling about some of them, but very few,” Sive says.

Sive and her postdocs began by identifying zebrafish genes analogous to the human genes found in this region. (Zebrafish, those genes are not clustered in a single genetic chunk, but are scattered across many chromosomes.) The researchers studied one gene at a time, silenc- ing that with short strands of nucleic acids that target a particular gene and prevent its protein from being produced.

For 21 of the genes, silencing led to abnormal development. Most produced brain defects, including improper development of the brain or eyes, thinning of the brain, or inflation of the brain ventricles, cavities that contain cerebrospinal fluid. The researchers also found abnormalities in the wiring of axons, the long neural projections that carry messages to other neurons, and in simple behaviors of the fish. The results showed that the zp1261 genes are very important during brain development, helping to explain the connection between this region and brain disorders.

Furthermore, the researchers were able to re- store normal development by treating the fish with the human equivalents of the genes that had been repressed.

“That allows you to deduce that what you’re learning in fish corresponds to what that gene is doing in humans. The human gene and the fish gene are very similar,” Sive says.

To figure out which of these genes might have a strong effect in autism or other disorders, the researchers set out to identify genes that produce abnormal development when their activity is reduced by 50 percent, which would happen in someone who is missing one copy of the gene. (This correlation is not seen for most genes, because there are many other checks and balances that regulate how much of a particular protein is made.)

The researchers identified two such genes in the zp1261 region. One, called kif7, codes for a protein involved in the separation of chro- mosomes during cell division, and one, aldolase a, is involved in glycolysis—the process of breaking down sugar to generate energy for the cell.

In work that has just begun, Sive’s lab is work- ing with Stanford University researchers to explore in mouse predictions made from the zebrafish study. They are also conducting molecu- lar studies in zebrafish of the pathways affected by these genes, to get a better idea of how defects in these might bring about neuro- logical disorders.

FISH ON THE BRAIN

In zebrafish embryos, interrupting the function of the gene kif7 causes the retina (dotted arrow) to develop abnormally, indicating that this gene is one of many vital for proper development. Courtesy Denise Melnick & Michael Blevins
The kids are up to their old tricks on this dazzling late winter afternoon. Bryson and Sanchez move in lock step, one rarely out of touch with the other. Milo, the certified bad boy of the bunch, sends a swift but playful hind-leg kick toward one of his handlers, who’s having none of it; and Randance, the reluctant alpha, seems a tad puzzled when the others fall in behind during his slow stroll across the yard.

And as they go about the fairly mellow business of being alpacas, they’re all blissfully unaware that they are participants in a set of grand experiments with the potential to forever alter the landscape of biomedical research.

Among the drivers of these experiments is Whitehead Institute Member Hidde Ploegh, who has begun to leverage one remarkable aspect of the alpaca immune system. It turns out that alpacas are capable of generating two kinds of antibodies: the traditional two-chain (heavy and light chains) antibodies found in mammals and vertebrates, and smaller, single-chain antibodies (heavy chain only).

First reported in 1993 by Belgian scientist Raymond Hamers, and further exploited by colleague Serge Muyldermans, these unique antibodies are produced by all members of the camelid family, including llamas, Bactrian camels, dromedaries, and, of course, alpacas.

This surprising discovery fueled a number of intriguing strategies for novel antibody engineering and therapeutic development, most of which rely on the use of a tiny, high-affinity domain of the antibody known as VHH. Although studied for nearly two decades, these single-domain or VHH antibodies hadn’t really captured Ploegh’s imagination until roughly three years ago as he sat in a lecture hall in Belgium for a talk by Jan Steyaert, an expert on camelid VHH production.

Suddenly, it all started to click. Ploegh realized that their small size—roughly 1/10th that of normal antibodies—and thermal stability made VHHs ideal for targeting antigens that reside inside cells of interest. Perhaps equally relevant, he recognized that VHHs would be amenable to modification via “sortagging,” a highly specific protein labeling technique developed in his lab, to track their activity within cells and to identify which antigen a single VHH binds. Ploegh figured that he might be able to immunize camelids with selected cellular material, have them generate VHHs in response, and use the VHH antibodies to probe the inner workings of cells with heretofore unachievable precision.

“I felt that a huge advantage of the single-domain antibodies is that they could easily be produced in bacteria, shared as plasmids, and used to create fusions with other proteins to endow them with new functions,” says Ploegh.

These unsuspecting alpacas just may hold the key to unlocking a host of cellular secrets.
produced in bacteria, shared as plasmids, and used to create fusions with other proteins to endow them with new functions,” says Ploegh. He returned to Cambridge energized and in search of cameldids, and in a bit of geographic good fortune, found some about an hour away. At the Cummings School for Veterinary Medicine at Tufts University in North Grafton, Massachusetts, Professor Chuck Shoemaker had been using alpacas to generate VHHs for his research in developing therapeutic antibodies.

“He just rang me up,” recalls Shoemaker. “Hidde reached out to me to discuss development of an alpaca model. Initially, he wondered if he could just get some immunized alpaca genetic material he could use to identify binding agents.”

Although the proposed research strategy was quite different from his own—“He’s intracellular, I’m extracellular,” Shoemaker jokes—he was happy to help Ploegh get started. These alpacas represent an extraordinary, multifaceted resource.

“We use them to identify and bind neutralizing toxins and viruses for therapeutic applications,” Shoemaker notes. “But it doesn’t have to be limited to that. And it shouldn’t. We have the unique ability to have alpacas here, grazing within a few hundred yards of the office. It’s a perfect situation to exploit this powerful tool.”

With help from Shoemaker and the Tufts Veterinary staff, Ploegh and his lab validated the concept by immunizing an alpaca with mouse spleen cells and harvesting VHH antibodies from a simple blood draw several weeks later. As hoped, the lab identified a VHH reagent specific for a subpopulation of mouse B cells. Ploegh described this proof-of-principle experiment in a special issue application that outlined his broader vision for a VHH antibody platform. Ultimately, he intends to develop large libraries of VHH antibodies that can bind and disrupt intracellular proteins in a variety of model organisms—from yeast to Drosophila. “This really does allow us to engineer the smaller antibodies we need. It works so beautifully, it’s pretty unbelievable,” says Cragnolini. “This really does allow us to engineer the smaller antibodies we need. It works so beautifully, it’s pretty unbelievable.”

Ingram has been working to help colleagues at MIT develop antibodies to yeast nuclear pore proteins, which could further enhance our understanding of one of the most closely studied organisms on the planet. She marvels at the efficiency of the approach.

“We can use such reagents within six months of the first alpaca immunization,” Ingram says. “We can conduct important experiments right away.”

Meanwhile, back in North Grafton, where strict protocols protect the animals from harm, it would seem the only real risks facing the alpacas are the inconvenience of the occasional photoshoot and perhaps a little frowning from visitors enthralled with these fascinating creatures. Fortunately, they’re willing to tolerate the attention in their distinguished, if unwitting, service to science.

“Intrigued? Please send your questions, comments, and suggestions to info@wi.mit.edu.

For his part, Ploegh remains convinced of the enormous benefits to science associated with large-scale development of VHH antibody libraries, despite the fact that his Pioneer Award from NHGRI is meant to fund “high risk, high reward” research.

“I think that, if properly developed, this approach may be an interesting complement to genetic techniques such as RNAi,” Ploegh asserts. “But time will tell—that’s why it’s a high risk, potentially high gain situation.”

In the meantime, as the Ploegh lab expands its efforts to engineer antibodies from a number of other species, including fish, flies, and worms, many are hopeful that the alpacas will continue to provide a rich resource for illuminating complex biological processes.

WANT TO GET INVOLVED?

Because Ploegh believes the applications for identifying VHH antibody targets are virtually limitless, and because he envisions this platform on a grand scale, he’s inviting inquiries and suggestions from all corners.

“For those who find this work appealing and who might be interested in learning more or even supporting an immunization schedule for an alpaca, we’d love to discuss the possibilities,” Ploegh says.

“For the scientists out there—be they biologists, chemists, biological or chemical engineers—we’d love to hear their ideas about where they think we should take this work,” he adds. “How do they see this approach being useful in their own endeavors? Such interactions could prove quite fruitful.”
According to Lewitter, BLAST’s expansion reflects the blossoming of the bioinformatics field. With the advent of inexpensive genome sequencing, numerous assays for RNAs and proteins, and genome-wide association scans, biology has created a flood of data. The exercise not only familiarizes students with the BLAST tool, but also reinforces the relationships between genes, proteins, and traits. According to Form, the ability to use bioinformatics to teach evolution, phylogeny, and even physiology makes it a very powerful teaching tool.

“It’s a great way for kids to visualize biology, and it got them asking their own questions, which is what scientists do,” says Form. “Some of the kids got so into it that I need to kick them out of the computer lab.”

Because BLAST is available online, it is apart of a scientist’s toolkit that high schools can provide with minimal investment. Sylvie Therrien, who is teaching AP and 10th grade biology at Dennis-Yarmouth Regional High School, has limited funds for wet lab supplies, so performing DNA extractions or other common lab tasks is not possible. With access to a computer lab, Therrien’s students can experience how scientists analyze data and reach conclusions. And her students have learned an important scientific truth: “Bad data is bad data.” That is, the quality of a conclusion depends on the quality of the original data.

Sometimes, however, the little things, such as a momentary delay, can have the biggest impact. “I think the students were frustrated that BLAST was taking a few seconds longer to respond to their requests. When Therrien explained that the delay was likely due to increases in usage by researchers returning from the holidays, her students got really excited.”

“It gave them the impression that they were accessing something big and important, something almost mysterious,” recalls Therrien. “It made it real, that BLAST is something that’s actually used and something they could use in real life, not just in some exercise.”

That reality may be closer than Therrien’s students realize. Several of Form’s former students have started bioinformatics careers in college, and one recently received his doctorate in the subject.

As Lewitter assesses the state of her program, she knows the future is bright for students entering the field—the demand for professionals who know how to store, analyze, and understand the enormous data pouring out of labs around the world will only increase.

There’s a growing need for new people to enter the field with a lot of jobs in industry and academia,” she says. “High school students are at the point when they’re trying to figure out what direction they want to go in. Getting them exposed to bioinformatics in class is important; up those young minds know it’s an option to pursue this area.

During this one-day workshop, which Amy Tremblay from Whitehead’s Department of Communications helped organize, Fran Lewitter, Director of Whitehead’s Bioinformatics and Research Computing (BaRC), gave Del Little, a science teacher at Windsor School in Boston, a few observations about BLAST’s collaborative nature. Because BLAST is a free, online tool that scientists can use to compare genetic sequences or proteins, and develop functional or evolutionary relationships between various organisms, it is an important tool for integrating bioinformatics into the classroom. According to him, the committee was charged with increasing the enrollment in introductory bioinformatics exercises, including one that uses BLAST to produce a phylogenetic tree for the Plasmodium parasite that causes malaria.

Brad Williamson, a Master Teacher in the University of Kansas UKTeach Program and a member of the AP Biology Development Committee, that revamped the curriculum. According to him, the committee was charged with rethinking the depth and context of understanding across all of biology and introducing to students the process of thinking like a scientist.

According to Lewitter, BLAST’s expansion reflects the blossoming of the bioinformatics field. With the advent of inexpensive genome sequencing, numerous assays for RNAs and proteins, and genome-wide association scans, biology has created a flood of data. The exercise not only familiarizes students with the BLAST tool, but also reinforces the relationships between genes, proteins, and traits. According to Form, the ability to use bioinformatics to teach evolution, phylogeny, and even physiology makes it a very powerful teaching tool.

“It’s a great way for kids to visualize biology, and it got them asking their own questions, which is what scientists do,” says Form. “Some of the kids got so into it that I need to kick them out of the computer lab.”

Because BLAST is available online, it is apart of a scientist’s toolkit that high schools can provide with minimal investment. Sylvie Therrien, who is teaching AP and 10th grade biology at Dennis-Yarmouth Regional High School, has limited funds for wet lab supplies, so performing DNA extractions or other common lab tasks is not possible. With access to a computer lab, Therrien’s students can experience how scientists analyze data and reach conclusions. And her students have learned an important scientific truth: “Bad data is bad data.” That is, the quality of a conclusion depends on the quality of the original data.

Sometimes, however, the little things, such as a momentary delay, can have the biggest impact. “I think the students were frustrated that BLAST was taking a few seconds longer to respond to their requests. When Therrien explained that the delay was likely due to increases in usage by researchers returning from the holidays, her students got really excited.”

“It gave them the impression that they were accessing something big and important, something almost mysterious,” recalls Therrien. “It made it real, that BLAST is something that’s actually used and something they could use in real life, not just in some exercise.”

That reality may be closer than Therrien’s students realize. Several of Form’s former students have started bioinformatics careers in college, and one recently received his doctorate in the subject.

As Lewitter assesses the state of her program, she knows the future is bright for students entering the field—the demand for professionals who know how to store, analyze, and understand the enormous data pouring out of labs around the world will only increase.

There’s a growing need for new people to enter the field with a lot of jobs in industry and academia,” she says. “High school students are at the point when they’re trying to figure out what direction they want to go in. Getting them exposed to bioinformatics in class is important; up those young minds know it’s an option to pursue this area.

During this one-day workshop, which Amy Tremblay from Whitehead’s Department of Communications helped organize, Fran Lewitter, Director of Whitehead’s Bioinformatics and Research Computing (BaRC), gave Del Little, a science teacher at Windsor School in Boston, a few observations about BLAST’s collaborative nature. Because BLAST is a free, online tool that scientists can use to compare genetic sequences or proteins, and develop functional or evolutionary relationships between various organisms, it is an important tool for integrating bioinformatics into the classroom. According to him, the committee was charged with increasing the enrollment in introductory bioinformatics exercises, including one that uses BLAST to produce a phylogenetic tree for the Plasmodium parasite that causes malaria.

Brad Williamson, a Master Teacher in the University of Kansas UKTeach Program and a member of the AP Biology Development Committee, that revamped the curriculum. According to him, the committee was charged with rethinking the depth and context of understanding across all of biology and introducing to students the process of thinking like a scientist.
I couldn’t be that easy. Whitehead Institute researchers couldn’t possibly identify anonymous participants in genomic studies simply by using a computer, an Internet connection, and publicly accessible online resources. Could they?

Indeed they could. And they did. In what began as an exercise in “vulnerability research”—a common practice in the field of information security—Whitehead Fellow Yaniv Erlich and Melissa Gymrek, a student in Erlich’s lab, took a multi-step approach to prove that under certain circumstances, the full names and identities of genomic research participants can be determined, even when their genetic information is held in databases in de-identified form.

In a paper that has fueled a global debate about privacy protection, Erlich and Gymrek reported earlier this year in the journal Science that their methods enabled them to identify nearly 50 individuals who had submitted personal genetic material as participants in genomic studies.

“This was an important result that served to point out the potential for breaches of privacy in genomics studies,” says Erlich. “The threat is very real.”

Erlich and colleagues began by analyzing unique genetic markers known as short tandem repeats on the Y chromosome (Y-STRs) of men whose genetic material was collected by the Center for the Study of Human Polymorphisms (CEPH) and whose genomes were sequenced and made publicly available as part of the 1000 Genomes Project. Because the Y chromosome is transmitted from father to son, as are family surnames, there is a strong correlation between surnames and the DNA on the Y chromosome. Recognizing this correlation, genealogists and genetic genealogy companies have established publicly accessible databases that house Y-STR data by surname. In a process known as “surname inference,” the Erlich team was able to discover the family names of the men by submitting their Y-STRs to these databases. With surnames in hand, the team queried other information sources, including Internet search engines, obituaries, genealogical websites, and public demographic data from the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at New Jersey’s Coriell Institute, to identify nearly 50 men and women in the United States who were CEPH participants.

Previous studies have contemplated the possibility of genetic identification by matching the DNA of a single person, assuming the person’s DNA was cataloged in two separate databases. This work, however, exploits data between distant paternal relatives and not merely between a single individual and any database. As a result, Erlich and Gymrek say that the posting of genetic data from a single individual can reveal deep genealogical ties and lead to the identification of a distance-related person who may have no acquaintance with the person who released that genetic data.

“We show that it, for example, your Uncle Dave submitted his DNA to a genetic genealogy database, you could be identified,” says Gymrek, first author of the Science paper. “In fact, even your fourth cousin Patrick, whom you’ve never met, could identify you if his DNA is in the database, as long as he is paternally related to you!”

Aware of the sensitivity of his work, Erlich emphasizes that he has no intention of revealing the names of those identified, nor does he wish to see public sharing of genetic information curtailed.

“Our aim is to better illuminate the current status of identifiability of genetic data,” he says. “More knowledge empowers participants to weigh the risks and benefits and make more informed decisions when considering whether to share their own data. We also hope that this study will eventually result in better security algorithms, better policy guidelines, and better legislation to help mitigate some of the risks described.”

To that end, Erlich shared his findings with officials at the National Human Genome Research Institute (NHGRI) and NIGMS prior to publication. In response, NIGMS and NHGRI moved certain demographic information from the publicly-accessible portion of the NIGMS cell repository to help reduce the risk of future breaches. In the same issue of Science in which the Erlich study appears, Judith H. Greenberg and Eric D. Green, the Directors of NHGRI and NIGMS, and colleagues authored a perspective on this latest research in which they advocate for an examination of approaches to balance research participants’ privacy rights with the societal benefits to be realized from the sharing of biomedical research data.

“This is not shocking; I think this is just a moment of recognition, a reflection moment,” Green told the Boston Globe. “We have all these values which are totally laudable, but are beginning to come into conflict. What is the best way to navigate this?”

Suggestions for successful navigation are myriad and are being proposed from all corners. An editorial in the influential journal Nature—which stated that “…Erlich responded in an exemplary way to his team’s findings by contacting the NIH and other gene- ics researchers with his findings before publishing them…”—declared that Erlich’s work “sets an important precedent for con- structively dealing with newly discovered privacy loopholes, and other researchers should take note."

“Yaniv’s work is a timely reminder that in this era in which massive amounts of genomic data are being generated rapidly and shared in the interest of scientific advancement, there is an increasing likelihood of privacy breaches,” says Whitehead Institute Director David Page. “I’m delighted that, thanks to Yaniv’s overture to NIH, we at Whitehead Institute have the opportunity to join policymakers at NHGRI and elsewhere in what will be a critical, ongoing dialogue about the importance of safeguarding data, of sharing data, and the implications of failure in either endeavor.”

More knowledge empowers participants to weigh the risks and benefits and make more informed decisions when considering whether to share their own data...
FEELING THE HEAT Whitehead Fellow David Pincus’s new lab investigates how the heat-shock response is regulated in cells, finding changes in their environment.

**David Pincus Investigates How Cells Strive for Stable Conformations Despite Adversity**

New Whitehead Fellow David Pincus, who arrived at the institute last September, is certainly not the first in the renowned Fellows Program to follow an unconventional path to biology. All too often, this pursuit is far too long and arduous. Realizing he had an aptitude for art and a love of the natural world, Pincus majored in English at the University of California, Berkeley, before transitioning to the life sciences.

In Pincus’s case, he was happily pursuing his passion, completing a dual undergraduate degree in the arts (including drawing, painting, and printmaking) and science (such as biochemistry, cell biology, and molecular biology). He has his sights set on sharing his scientific passion with the world.

“At Whitehead, Pincus has found a near-ideal environment in which to explore the complex relationship between yeast genotypes and phenotypes,” says Pincus. “The collaborative environment here is unparalleled.”

“SEBASTIAN LOURIDO HAS HIS SIGHTS SET ON PARASITES”

Sebastian Lourido wields the diverse, refus ing to be constrained by convention. It’s why this native of Cali, Colombia, spent one of his high school years in Germany, and why he chose to attend college in Baltimore at Tulane University. It’s why he completed a dual undergraduate degree in cellular and molecular biology and studio arts (including drawing, painting, and printmaking). And it’s why a graduate student at Washington University in St. Louis, and now at Whitehead’s newest Fellow, he has eschewed certain facets of traditional biology to delve into parasitology.

“Tissue tropism is unique, and it’s wildly suc cessful,” Lourido says. “Roughly 20% of the world’s population is infected with it, and yet the majority of those infected are asymptomatic.” Understanding Tissue tropism can tell us a lot about how parasites subvert the immune system and could contribute to our understand ing of malaria and other apicomplexan diseases. Lourido arrived at Whitehead last November and says that thus far, the Institute has exceeded his expectations. He’s benefiting from the unique combination of independence and support that is a hallmark of the Fellows program and is relishing the opportunity to explore a relatively unknown area of biology that is also medically significant.

“Integrating a more detailed understanding of these parasites is far too long and arduous. Realizing he had a love of the natural world, Pincus majored in English at the University of California, Berkeley, before transitioning to the life sciences. In Pincus’s case, he was happily pursuing his passion, completing a dual undergraduate degree in the arts (including drawing, painting, and printmaking) and science (such as biochemistry, cell biology, and molecular biology). He has his sights set on sharing his scientific passion with the world. At Whitehead, Pincus has found a near-ideal environment in which to explore the complex relationship between yeast genotypes and phenotypes,” says Pincus. “The collaborative environment here is unparalleled.”

**BASICALLY CLINICAL**

Sebastian Lourido has been a Whitehead Fellow since April 2014 at the Institute.

**Fink Becomes President-elect of AAAS**

Whitehead Institute Founding Member Gerald Fink has been chosen President-elect of the American Association for the Advancement of Science (AAAS). With his election, Fink began a three-year term as an officer and member of the Executive Committee of the AAAS Board of Directors and will assume the role of President of AAAS in February 2015.

“It’s an honor to be elected to the presi dency of AAAS, which serves as the voice of American science to the world,” says Fink, who served as Whitehead Institute Director from 1990 to 2001. “I am prepared to tackle a number of daunting challenges facing us today, not least of which is the waning support for American research by the Federal govern ment. As the eventual leader of AAAS, I intend to work hard to protect our most valuable group in science—today’s students, fellows, and young faculty—and to ensure that the world’s population has the benefit of their insights and ideas in the decades ahead.”

Fink, who is also a professor of biology at MIT, is a pioneering researcher in the field of yeast molecular biology. His discovery of Pcp1p, a central protein kinase (PKD2) in the lifecycle of Toxoplasma, paved the way for advances in basic science and the development of novel pharmaceuticals. He has also been active in science policy, including his work with the 2002 National Academy of Sciences committee that advised the Federal government on how to meet the threat of bioterrorism without jeopardizing scientific progress. A graduate of Amherst College, Fink received his PhD in genetics from Yale University. He is a past president of the Genetics Society of America, a member of the National Academy of Sciences, the Institute of Medicine, and the American Philosophical Society, and a fellow of the American Academy of Arts and Sciences.
FEELING INSECURE A recent publication from a team of Whitehead Institute researchers has raised a host of concerns about the security of personal genetic information housed in large online databases accessed by researchers around the world. To learn more about potential data vulnerabilities and their implications, turn to page 14.