



"Think of every fish as a gene," says Jason Moore of the image on the screen behind him. He works to develop ways to examine genetic data visually instead of on a spreadsheet.

Text by Amos Esty • Photographs by Jon Gilbert Fox

Moore and more

Jason Moore holds a paradoxical view of genetics research. On the one hand, he believes the field has become too entranced by technology. On the other hand, Moore is no Luddite. To make sense of genetic data, he uses supercomputers, innovative software, and even 3D imaging that replaces spreadsheets with real-life metaphors—such as the fish visible in the image on the facing page. Still, when he uses this cutting-edge technology, he does so to investigate a very old idea, one that he thinks has been overlooked in recent years.

Moore is among the minority of geneticists who have started to question the direction the field has taken since the completion of the Human Genome Project. During the 1990s, as researchers raced to produce a complete sequence of the human genome (the collection of DNA that makes up an individual's genetic inheritance), the implication was that breakthroughs in clinical genetics were only a matter of time.

By 2000, when a first draft of the sequence was published, the field had some very big expectations to meet. "It's impossible to overstate the significance of this achievement," *Time* magazine declared upon that occasion. "Armed with the genetic code, scientists can now start teasing out the secrets of human health and disease at the molecular level—secrets that will lead at the very least to a revolution in diagnosing and treating everything from Alzheimer's to heart disease to cancer, and much

Computational geneticist Jason Moore believes that more is needed to parse the human genome than simply more data. He has devised new ways to visualize genetic patterns. He built a supercomputer to crunch statistics faster. He tweets and blogs regularly about his work. And he ponders deeply on the insights of his scientific forebears.

For a **WEB EXTRA** with links to Moore's website, blog, and Twitter feed, see dartmed.dartmouth.edu/sp10/we05.

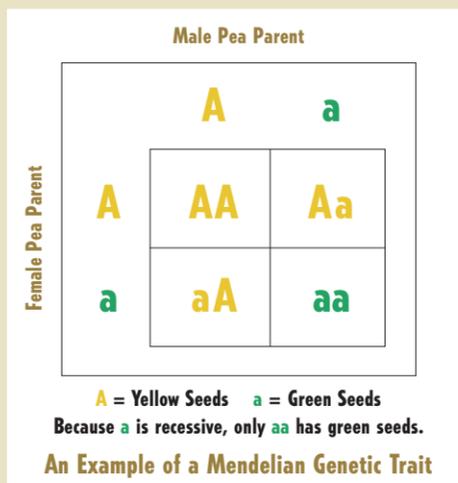
Amos Esty is the managing editor of DARTMOUTH MEDICINE.

One reason it's difficult to find genetic variations that explain disease risk is that there's simply so much information available. The human genome consists of about three billion pairs of nucleotides—the chemicals that make up DNA.

A look at Mendelian and epistatic traits in plants

In the 1860s, Gregor Mendel laid the foundation for modern genetics with his studies of inheritance in pea plants. The laws he developed predict traits that are determined by a single gene (which he called a “factor”). He found, for example, that the color of pea seeds is determined by just one gene. One variant of the gene produces yellow seeds, the other one, a recessive variant, produces green seeds.

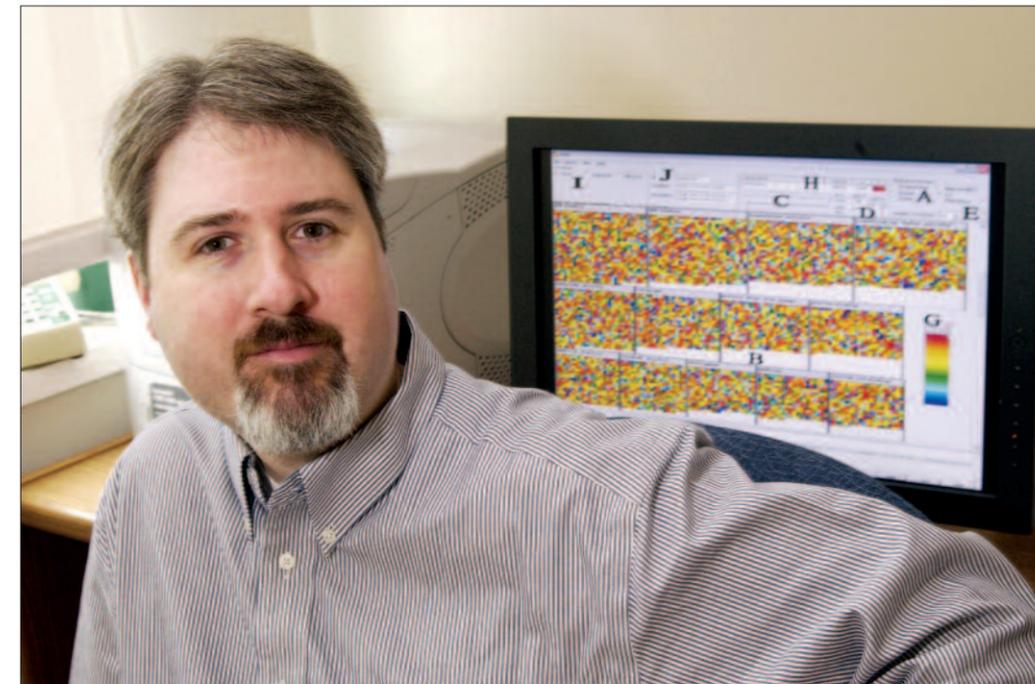
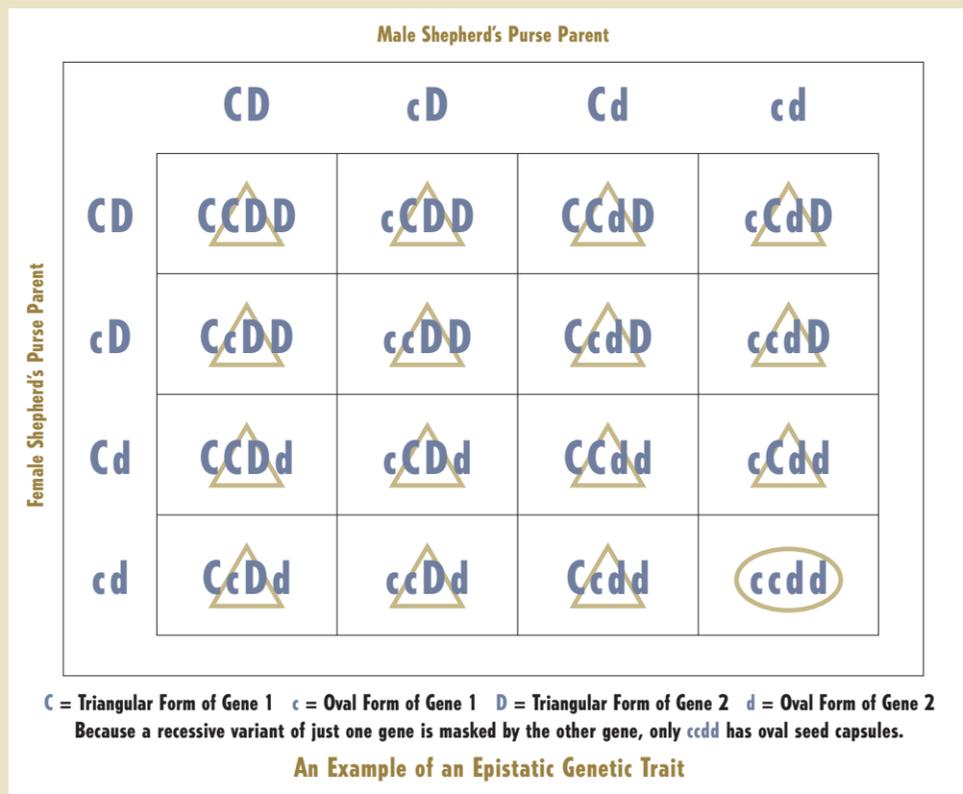
So if two parent plants that each carry one copy of the yellow variant and one copy of the green variant are crossed, the next generation will have roughly three plants with yellow seeds for every one plant with green seeds—as illustrated in the chart to the right.



By contrast, epistasis describes genetic traits that do not follow the laws of Mendelian inheritance because of interactions among multiple genes. One example was found in the early 20th century, when George Shull did experiments with a plant called shepherd's purse.

The shape of the plant's seed capsules—either triangular or oval—is determined by two

genes. For a plant to produce oval capsules, it must inherit two copies of a recessive form of two different genes. If parent plants that have one copy of the recessive form of each gene are crossed, the ratio of triangular to oval seed capsules in the next generation will be 15 to 1, as illustrated in the chart below.



Moore, pictured here shortly after his 2004 arrival at Dartmouth from Vanderbilt, began right away collaborating with researchers from other disciplines—cancer biology, epidemiology, and the neurosciences, to name just a few.

Cystic fibrosis, for example, is caused when a child inherits a recessive version of one gene from both parents. But only a small percentage of people carry even one copy of that gene, so the odds of getting cystic fibrosis are slight.

more. In a matter of decades, the world of medicine will be utterly transformed.”

Only a single decade has passed, but some scientists, including Moore, are now asking whether those promises can be kept. “The Human Genome Project was way overhyped with respect to at least its short-term health implications,” Moore says. “The sales pitch was that if we know the DNA sequence we’re going to cure diseases. Well, that didn’t really happen.”

The world is still awaiting major breakthroughs based on the Human Genome Project, points out Moore, who holds a Ph.D. in human genetics from the University of Michigan. Geneticists know that an individual’s specific DNA sequence harbors clues to predicting whether that person is particularly susceptible to one disease and not another. But so far, even knowing the complete sequence of the human genome, researchers haven’t been able to identify which genetic variants are important for most common diseases. Some scientists have even started referring to this missing genetic variability as the “dark matter” of genetics—they know it exists, but they can’t see it.

One reason it’s difficult to find genetic variations that might explain disease risk is that there’s simply so much information available. The human genome consists of about three billion pairs of nucleotides—the chemicals that make up DNA. The four possible nucleotides are represented by the let-

ters A, T, C, and G, and humans share the same sequence of letters on about 99.9% of the human genome. But that still leaves millions of locations where, for example, one person might have an “A” and another might have a “G.” A location where different people have different DNA letters is called a single nucleotide polymorphism (or SNP, pronounced “snip”). To try to predict a person’s risk of disease, scientists pore over SNPs, searching for variations associated with a particular disease.

This approach to identifying genetic risk factors grew out of the success scientists had finding genes that cause rare diseases such as cystic fibrosis and Huntington’s disease. Such diseases—which geneticists refer to as Mendelian diseases (see the box on page 28 for more on this subject)—are easily predicted by looking at single genes. Cystic fibrosis, for example, is caused when a child inherits a recessive version of one gene from both parents. But only a small percentage of people carry even one copy of that gene, so the odds of getting cystic fibrosis are slight.

With the completion of the Human Genome Project, geneticists had hoped to be able to find similarly important genes for more common diseases, such as cancers and heart disease. For the past five years, a common approach has been the use of genome-wide association studies. In these studies, scientists look at up to a million spots on the



Moore also mentors graduate students in Dartmouth's Program in Experimental and Molecular Medicine. Here, he confers with sixth-year Ph.D. student Kristine Pattin, who studies how information about the interaction of proteins can be used to improve disease prediction in genome-wide association studies.

"If those common variants had big effects on disease, we would all be sick," says Moore. "We would all have cancer. We would all have heart disease." So he focuses his research on the concept of epistasis, the interaction of genes.

genome, comparing the genetic sequences of people with the disease to the sequences of people without the disease. The goal is to identify single variations—SNPs—that are associated with the disease. That is, if at a specific point in the genome, most people with, say, lung cancer had an "A," but most people without lung cancer had a "G," then it could be an indication that that particular SNP was a predictor of the disease.

But the results so far have been disappointing. The few associations that have been found in one study usually have not been replicated in follow-up studies. And the associations that have been replicated usually explain only a small fraction of the cases of a disease. "That approach has not really worked very well at all," Moore says. "Common diseases are much more complex than Mendelian diseases, and it's not going to be single genes that predict with any certainty whether someone is going to develop disease."

As improving technologies allow scientists to look at more and more SNPs, it's possible that the "missing heritability" will be found. But Moore doesn't seem surprised that this approach hasn't yet

revealed more of the secrets of disease prediction. When the missing heritability is located, he argues, it will be the result of taking a more intelligent approach to the data, rather than just accumulating endless lists of SNPs.

One problem with these studies, he says, is that they overlook the complexity of genetic interactions. He thinks it's more likely that a number of genetic variations together determine disease risk. He points out that if common diseases were the product of single genetic mutations, it's likely that those mutations would have been weeded out from the human population by natural selection. "If those common variants had big effects on disease, we would all be sick," he says. "We would all have cancer. We would all have heart disease." So Moore's research focuses on the concept of epistasis, the interaction of genes.

"Epistasis is an old idea," Moore says. Indeed, in 1909, early geneticist William Bateson coined the term "epistatic" to describe the way one gene could mask the effects of another gene. At the time, Moore says, epistasis was used to describe genetic phenomena that didn't fit the results that would be expected from classical Mendelian genetics. Today,

epistasis is often used more broadly to refer to the interactions between different genes. "It's becoming a more popular idea, that genes don't work in isolation," Moore says. "Genes work together as part of large interactive networks."

Moore's longtime collaborator Scott Williams, Ph.D., a professor of genetics at Vanderbilt, has worked with Moore on honing this approach. "Genetic association studies haven't delivered on their promise," Williams says. "And the reason, we think, is that epistasis is so prominent."

In a recent article, Moore argued that disease can be viewed as a collection of different genetic variations that together overcome the body's ability to maintain a healthy state. So, rather than look for single common variations, it makes more sense to search for combinations of variations—that is, for epistasis. "I think the problem with the current approach is it assumes there's a simple relationship between the gene and the disease," he says. "We know heart disease is complex—just because you smoke doesn't mean you're going to get heart disease. There are no silver bullets that predict with certainty that you're going to get cancer or heart disease."

But the more combinations a researcher looks at, the more computing power the research requires. "We're talking about massive amounts of data," Moore says. "When you have a million genetic variations in your study, it turns out there aren't enough computers in the world to enumerate all the three-way and four-way combinations."

While he was an assistant professor at Vanderbilt, Moore began developing software called Multifactor Dimensionality Reduction (MDR) to tackle this problem. He continued the work after coming to DMS in 2004, eventually producing an open-source software package that allows researchers to more easily examine genetic data for epistasis. An advantage of MDR is that it can take combinations of two or more factors and turn them into a single variable, enabling researchers to look at far more interactions than they otherwise could.

The software is freely available on Moore's website (www.epistasis.org/) and has been downloaded tens of thousands of times. "People are using it to study all sorts of different diseases," he says.

Moore himself has worked with researchers at DMS to examine genetic variants in a number of diseases. In one study, he and collaborators in the Section of Epidemiology looked for risk factors for bladder cancer. A typical genome-wide association study would have looked broadly at hundreds of thousands of spots along the genome to find any that stood out as predictors of the disease. But

rather than take that approach, the Dartmouth team focused closely on SNPs in a few genes known to be involved in repairing damaged DNA.

Cells are equipped with mechanisms that enable them to repair damaged DNA, but when mutations knock out that ability, the result can be uncontrolled cell division—otherwise known as cancer. So it makes sense, Moore says, to look specifically at genes that are known to have a connection to the condition being studied. Some of the genes that control DNA repair are redundant—if one stops working, another is there to ensure that the DNA still gets repaired. So it could take more than a mutation to a single gene to stop DNA repair from happening, meaning that a single mutation might not reveal much about the risk of disease.

The researchers used MDR to compare several genes in people with and without bladder cancer, as well as to look at environmental factors such as age and smoking. When they looked for the one factor that best predicted the risk of bladder cancer, they found, not surprisingly, that it was smoking. That meant smoking was a better predictor of whether an individual would get bladder cancer than any single genetic variable. None of the genetic variants by itself was a significant predictor of risk.

Then the researchers looked at combinations of two factors. The software analyzed the effects of every possible two-way combination of variations, plus the effects of every individual gene combined with smoking and the other environmental factors. This time, Moore says, the results were surprising. Individuals with two specific variants on one DNA repair gene had a significantly increased risk of bladder cancer. In fact, that specific combination of variations was an even more robust predictor of disease than was smoking alone. "That was an unexpected, very interesting, and important finding," Moore says. "What was interesting about the study was that the two genetic variations had an epistatic relationship. In other words, the two variations together in MDR did a much better job of predicting bladder cancer than either one did individually."

Scott Williams says an important use of MDR is to go back over data that has already been collected to look for missed connections. What's needed in genetics, he says, is not necessarily more data as much as better analysis of data that's already been gathered. He thinks there could be a number of connections not recognized because researchers haven't looked for epistasis. "I think there's a lot more out there than people have discovered using the data that they've already published," he says.

continued on page 56

"We're talking about massive amounts of data," Moore says. "When you have a million genetic variations in your study, it turns out there aren't enough computers in the world to enumerate all the three-way and four-way combinations."

Dr Thomas G Schell and Dr Patrick C Noble, PLLC
Family & Esthetic Dental Care



Gentle family dental care and appliances for
 the treatment of Obstructive Sleep Apnea

31 Old Etna Road, Suite N1
 Lebanon, New Hampshire (603) 448-3800
www.gentledentalcare.com

BRATTLEBORO, VT 802-258-3070 BURLINGTON, VT 802-863-2375 LEBANON, NH 603-448-2211 MONTPELIER, VT 802-225-5500 PLATTSBURGH, NY 518-561-0068 ST. JOHNSBURY, VT 802-748-8324

THE VALUE OF FOCUS



Estate Planning, Business Succession, and Trust and Probate Administration

DRM's estate planning attorneys provide practical, timely, sophisticated legal advice to help our clients plan and manage assets to achieve specific financial objectives.

- Estate Planning & Administration
- Probate Administration & Litigation
- Qualified Personal Residence Trusts
- Estate Planning in Uncertain Times
- A/B Trust Tax Savings Strategies
- Charitable Remainder Unitrusts
- Irrevocable Life Insurance Trusts
- Family Limited Liability Companies



8 South Park Street, Lebanon
 603-448-2211 www.drm.com

LEXIS • MUNDI
 THE WORLD'S LEADING PROVIDER OF INDEPENDENT LAW FIRMS

BUSINESS LAW • CAPTIVE INSURANCE • BANKRUPTCY & BUSINESS RESTRUCTURING • LABOR & EMPLOYMENT LAW
 TAX LAW • HEALTH LAW • ENERGY & TELECOMMUNICATIONS • INTELLECTUAL PROPERTY LAW • TRUSTS & ESTATES
 FAMILY LAW • REAL ESTATE, LAND USE & ENVIRONMENTAL LAW • GOVERNMENT & PUBLIC AFFAIRS • LITIGATION

Moore and more

continued from page 31

Even with MDR, it can take a lot of computer power to analyze genetic data. To help with the problem, Moore facilitated the creation at Dartmouth of a supercomputer called DISCOVERY (an approximate acronym for Dartmouth Initiative for SuperCOmputing Ventures in Education and Research). It is made up of more than 900 processors and can be accessed by researchers in departments all across Dartmouth.

"The downside of these methods is that they're very computationally intensive, so we really need that resource," Moore says. By making the supercomputer available to other departments, the entire campus benefits. "There's an economy of scale that's really valuable, so we don't each have to build our own systems and have our own infrastructure and personnel, which would be very redundant and costly," explains Moore.

In addition to developing more powerful methods of analyzing data, he is interested in finding new, useful, and even fun ways to present data. The goal, he says, is to get away from spreadsheets. "No scientist likes looking at thousands and thousands of numbers in an Excel spreadsheet," he says.

One of Moore's newest efforts in that regard is something that he calls a visualization lab. It's a room crowded with high-tech gear. A six-foot-wide screen sits at one end of the room. With two high-definition projectors, Moore can use the large screen to present genetic data in some unexpected ways. "We're visual creatures," Moore says. "We respond emotionally and physically to visual stimuli, and why not incorporate that into the scientific discovery process?"

In one scenario, genes are represented by fish. The color and size of the fish can represent different genetic variants. A researcher can don 3D glasses and move through the ocean of data searching for interesting patterns or unusual groupings. "Think of every fish as a gene," Moore says. "Every fish is a gene or a genetic variant, and the size, the shape, the color, the different features of the fish are representative of the values of those different analytical methods that you've used. So now you can interactively explore this space." That makes it much easier to identify patterns quickly, he says. "Because it's a video game, you can fly through them,

you can interact with the fish, you can move them around, you can sort them."

The lab relies on technology developed for the video game industry. Moore's team licensed several programs that are used to create video games and instead uses them to develop these tools. "It's a really fun project, and I really believe that's the future. These fancy algorithms are important, but they're not going to get us all the way there."

There are still more components Moore would like to add to the lab. "What I hope in a year is that we'll have a completely immersive and interactive environment that's easy to use and intuitive for people to make discoveries," he says. "The hope is that you'll be able to make discoveries through this interactive process or at least . . . generate some hypotheses that you would not have thought of looking at a big Excel spreadsheet."

All of Moore's projects, from creating software to visualizing data, are intended to help shape the direction of genetics research. "One of my objectives is to change the field," he says. "I really see that as one of my missions as a scientist."

To do so, he tries to spread the word in as many ways as possible. In addition to giving talks and writing papers, he has an active blog (compgen.blogspot.com) that covers his own work and genetics research generally, and he uses Twitter as well (under the user name "moorej"). "It's really a way to push ideas out to the [scientific] community and to remind people about things that I think are important and to help people think about the problem differently. It's a way for me to have an impact on the rest of the community."

Yet despite his use of new media and advanced technology, Moore spends a lot of time thinking about the past. "We have a lot to learn from early geneticists," he says. "They were smart people who were really thinking deeply about the problem."

Today, he argues, genetics students spend too much time learning to use the newest equipment and too little time reading the old genetics literature. Not surprisingly, given his ambivalent attitude toward technology, Moore believes in the importance of history. "Historical context is so important for what we do," he says. "It provides a grounding, a foundation. You have to understand the history in order . . . to understand your place in the science." ■

PARTNERS FOR LIFE



Adele and Hugh

Diagnosed with multiple sclerosis at age 21, Hugh Edgerton lived with the progressive disease for more than 60 years. Nonetheless, he and Adele, his wife of almost as many years, lived their life together to the fullest. "Hugh was one of those optimistic people who was confident that a cure will be found," says Adele.

It is that hope that inspired Hugh and Adele to establish a charitable gift annuity with DHMC, designating that their gift advance neurological

research. Funded with stock that had grown in value over many years, their gift provided Adele with a charitable income tax deduction and a fixed, guaranteed income for the rest of her life. "It seems like the perfect solution," says Adele.

CHARITABLE GIFT ANNUITY FEATURES

- guaranteed fixed income for life
- partially tax-free income
- charitable tax deduction
- cash or appreciated assets may be gifted
- income for one or two lives

SAMPLE RATES

Age	Rate
65	5.3%
70	5.7%
75	6.3%
80	7.1%
85	8.1%
90+	9.5%

Contact us today for your one-life or two-life rate.

Toll Free 1-866-272-1955

Office of Gift Planning
 E-Mail: Gift.Planning@Hitchcock.org
 Web site: www.dhmc.org/dept/dev/cga

The Power of Partnership

DARTMOUTH-HITCHCOCK
 MEDICAL CENTER

Dartmouth
 Medical School