



MICHAEL MCALEAR

MICHAEL WEIR

ROBERT LANE

LAURA GRABEL

BIOLOGY'S NEW DATA

Torrents of data have spilled out of the Human Genome Project, leading biologists to approach their science in an entirely new way. BY WILLIAM L. HOLDER '75

People are data.

That's the mantra for a new breed of biologists who use computers to sort through vast databases comprising billions of bits of information about genes. They are searching for patterns that reveal why one person is different from another, why one person's genes portend cancer while another's augur a long, healthy life.

The urgent need to make sense of the prodigious amounts of data produced by the Human Genome Project, as well as rapidly accruing information about non-human genomes, has brought biology and computer science down the aisle and created the science of bioinformatics. Almost overnight, biologists have discovered that the research problems they face require the expertise of individuals who may never have dissected a frog but have spent their careers peeling apart computer algorithms. Bioinformatic centers are sprouting up from Tehran to Texas.

Wesleyan's recently formed Integrated Genomic Sciences program is designed to place the university at

the forefront of bioinformatics. The program takes advantage of the cross-disciplinary connections that a small university can facilitate. Faculty from biology, molecular biology, chemistry, physics, computer science, and philosophy are participating in this effort to pool intellectual resources in ways rarely, if ever, seen before among Wesleyan scientists.

"Bioinformatics is affecting how all of us in the life sciences think about our research," says Laura Grabel, a biologist serving as dean of the sciences and mathematics. "Now you can look at the activity of every single gene in a cell during some transition, for example, from normal cell to cancer cell. You are going to turn up things you hadn't thought were relevant to your interests."

Michael McAlear, associate professor of molecular biology and biochemistry, adds: "You can't be a functioning life scientist without seeing bioinformatics in your face every week."

As he spoke, McAlear thrust forward the Oct. 4 issue of *Science* magazine, which contains analyses

of the genomes of the malarial mosquito and the parasite it carries. By using bioinformatics to study the 15,000 genes in the mosquito and the 5,000 genes in the parasite, scientists may be able to create new designer drugs to replace existing antimalarials that are succumbing to disease resistance. Alternatively, they may be able to genetically modify the mosquito itself so that it cannot host the parasite. More generally, scientists believe that bioinformatics will greatly speed up the pace of discoveries about genes and the development of new drugs.

"Experiments you never dreamed of are now doable with bioinformatics techniques," says McAlear. But then he adds a cautionary note: "The gene for cystic fibrosis was cloned and sequenced more than 10 years ago. We still don't have a cure."

If the genome were a city, the field of bioinformatics would include the tools and analytical approaches for taking a census of all the residents, learning who they are, what jobs they have, and where they relax. On a cellular level this is what bioinformatics makes

possible: the study of groups of genes, learning how they live and work together.

The significance of this ability can hardly be overstated. Biologists have spent enormous amounts of time laboring over genes, one by one. With bioinformatics, they can study thousands of genes at once, observing the intricacies of their interactions in ways never before possible.

Take the sense of smell, for instance. People can detect about 10,000 scents, ranging from the delight of freshly baked bread to the stench of an angry skunk. Smell fascinates Robert Lane, a new faculty member in molecular biology and biochemistry. With a doctorate from Caltech and having done his postdoctoral work with Leroy Hood of the Institute for Systems Biology in Seattle, Lane is the first scientist hired by Wesleyan who has been educated in an integrative genetics program. The university was able to hire him with funding from the Keck Foundation, which has supported the IGS program, along with the Howard Hughes Medical Institute and the President's Fund for Innovation at Wesleyan.

Lane appreciates the sense of smell not only for its exquisite versatility, but also for its adaptability. Unlike any other kind of neuron, the millions of neurons in the posterior of the nose that are responsible for initial odor detection regenerate every couple of months and are hypothesized to adapt according to an animal's needs or stage of life. No other developmental system is so critical to reproduction and survival yet so dynamic.

People are sensitive to a wide range of odors because we each have about 1,000 genes that carry the genetic code to create 1,000 different odor receptors active in the nasal neurons. Roughly three percent of human genes are devoted to smell, which underscores the importance of this sense.

Every olfactory neuron contains all 1,000 genes, yet only one is active per neuron. The implication, says Lane, is that the other 999 or so are silenced. How DNA conducts this orchestra of genes, calling on one while muting others, is Lane's principal area of research. The more scientists learn, the more they are coming to appreciate the complexity of this genetic symphony. For example, the location of a gene in its chromosome matters much more than was once imagined. The genes that code for odor receptors are located in "the darkest corners of the genome—something I've called transcriptional black holes," says Lane. He believes that all the olfactory genes are normally turned off and some-

how, one by one, are turned on to create individual odor receptors.

With bioinformatic approaches, Lane hopes to learn more about these "black hole" regions of the genome, to search for meaningful patterns in the dizzying amounts of data that make up the genetic code. His work has implications for gene therapy because it suggests that scientists will need to know exactly where to insert genes in order for them to work as expected.

A measure of Wesleyan's commitment to genomic sciences is its acquisition last spring of gene chip array technology, now coming into use at many large research universities. The array allows scientists to measure the activity of thousands of genes simultaneously. It works on the principle that an active gene churns out mRNA (messenger RNA), which, in turn,

A measure of Wesleyan's commitment to genomic sciences is its acquisition last spring of gene chip array technology, now coming into use at many large research universities.

guides the cell in the manufacture of each of the thousands of proteins it needs to survive and carry out its function. Scientists take a small glass chip, about the size of a postage stamp, that contains a grid of thousands of genes of known identity, then coat it with the mRNA of a cell of interest, say a cancer cell, plus a fluorescent dye. All the genes that correspond to the cancer mRNA fluoresce in a grid pattern that can be read by a computer.

The gene chip is the technological underpinning for the effort to see which genes are active in any given cell; that is, which are working together for the health or the detriment of the host.

Professor of Physics Rick Jensen, currently on sabbatical leave at Harvard Medical School, is analyzing gene chip data to look at differences between cancerous and noncancerous cells (work he is carrying out with Josh Blumenstock '03, coauthor of four papers with

Jensen). In the vast majority of diseases, afflicted cells have anomalous levels of various proteins. Scientists can use the chip array to see which genes in cancer cells are unusually active and correlate that information with protein levels. This quantitative analysis produces data sets with 40,000 numbers, and understanding them is next to impossible without computers.

"This is a natural playground for physicists," says Jensen.


Even before Wesleyan acquired its chip array, Jensen wanted to "have first crack at this new data," so he undertook an analysis of DNA chip information from Stanford researchers who were studying the genetic control of cell division in yeasts. He was looking for genes that start the process of cell division, and he found good candidates. When he checked the identity of the genes in a national database, he was surprised to discover that one, a gene known as EBP2, had been discovered by McAlear.

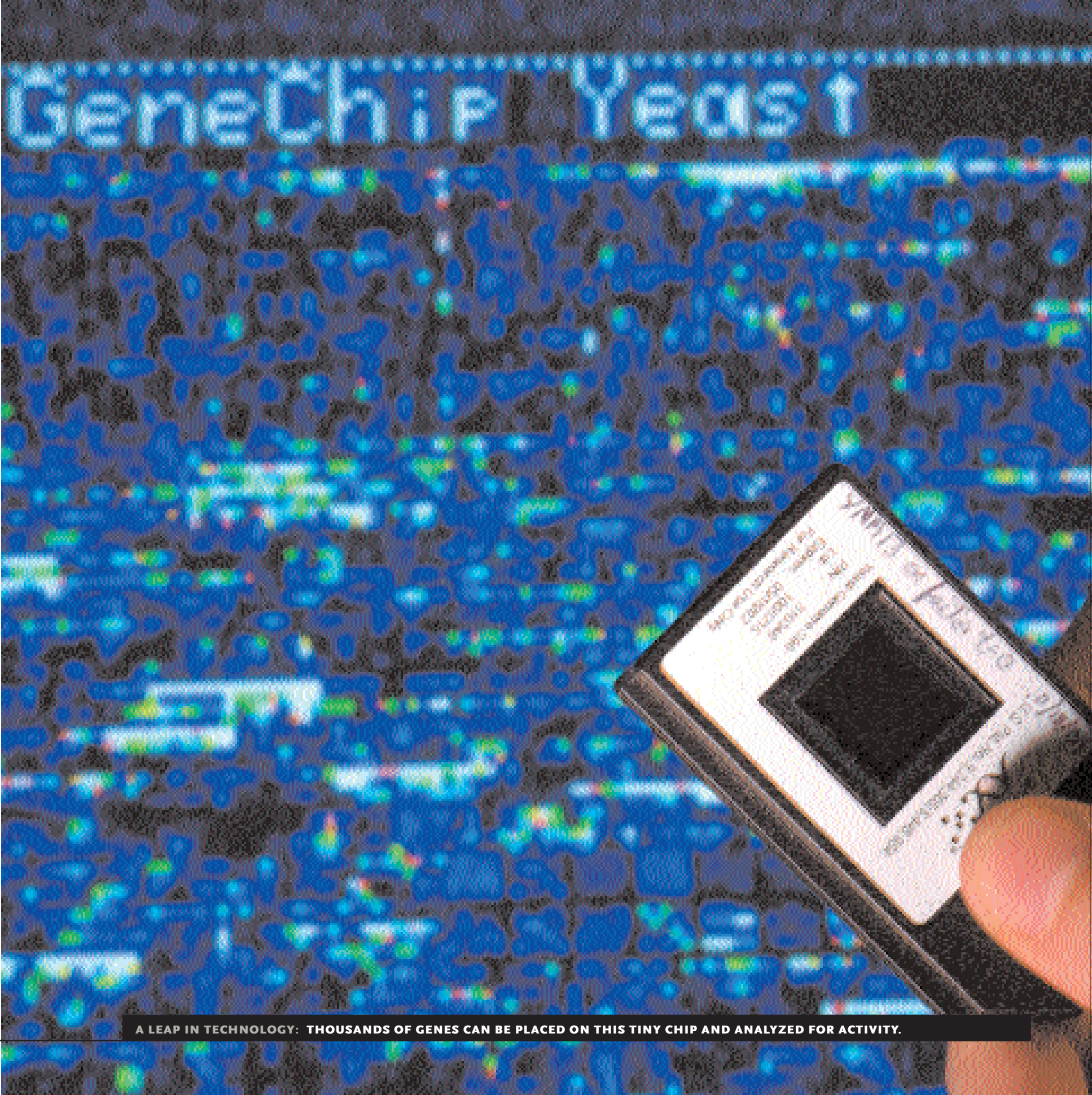
"So I told Mike about this. He was practically jumping up and down with excitement," Jensen recalls. No one knew that EBP2 might be involved in the initiation of cell division, and Jensen was able to suggest to McAlear that it works in concert with 60 other genes.

In the classroom, students are learning bioinformatics in a novel course that is team-taught by Michael Weir, professor of biology, and Michael Rice, professor of computer science. Weir emphasizes that their approach involves true team-teaching—both are present for every class, playing off each other, agreeing and disagreeing. Last year the class split equally between students majoring in life sciences and in computer science. Students collaborated on projects, sharing diverse backgrounds. The mix also presented a challenge to Weir and Rice, who could not make assumptions about background and familiarity with language or concepts.

"It's good when scientists don't assume they can use jargon," says Weir.

Students who study integrative genomics at Wesleyan will be entering a field with enormous promise. Decoding the human genome was only the first step in a monumental effort to understand which genes make which proteins, what these proteins do, and how the process can go wrong. Already a new word has been coined to describe this venture: proteomics. The need for biologists to understand how to analyze vast sets of data grows daily.

"The future of biology," says Lane, "is in informatic science." 



A LEAP IN TECHNOLOGY: THOUSANDS OF GENES CAN BE PLACED ON THIS TINY CHIP AND ANALYZED FOR ACTIVITY.