Science deals exclusively with things as they are in themselves.
John Ruskin (1819–1900)

TOXICOGENOMICS

Harnessing the HGP for Public Health

The enormous amount of genetic data from the Human Genome Project (HGP) has benefited researchers studying gene–environment interactions, but also presents challenges in experimental design, data management, and the ethical, legal, and social implications of gene–environment research. These challenges were the topic of the symposium "The Human Genome Project and Public Health: Gene–Environment Interactions," part of the 2004 annual meeting of the American Association for the Advancement of Science in Seattle, Washington.

Presenters discussed strategies to help researchers decide where to concentrate their research dollars and energy among the more than 7.2 million genetic variations and single-nucleotide polymorphisms (SNPs) cataloged by the HGP. One strategy is to focus on genes that are linked by function in pathways, said Deborah Nickerson, a professor in the University of Washington (UW) Department of Genome Sciences. She discussed progress on SNP discovery in gene pathways for the Environmental Genome Project, especially the development of new algorithms to explore associations between SNPs in human genes and environmental exposures. One program—Hotspotter, developed by UW assistant statistics professor Matthew Stephens—helps researchers explore recombination in human genes over time, a history that affects SNP associations in the human genome.

Focusing on genes involved in key cellular processes such as DNA repair is critical, according to symposium co-organizer David Eaton, director of the NIEHS–UW Center for Ecogenetics and Environmental Health. Every day, each cell in the human body withstands 10,000–20,000 oxidative assaults to DNA, and the vast majority are repaired. Increased understanding of how this DNA repair happens and other critical biochemical pathways can lead to better methods for disease prevention and more effective drugs with fewer side effects, said Eaton.

Better understanding of gene–environment interactions can also lead to more accurate dosing of existing medications. This is especially crucial for powerful drugs where the margin between effective and toxic doses is narrow, said Kenneth Thummel, associate dean for research and new initiatives of the UW School of Pharmacy. One example is warfarin, an anticoagulant used to prevent recurrent myocardial infarction and other thromboembolic events. Variations in a single gene (*CYP2C9*) can cause some people to be five times as sensitive to the drug and make them susceptible to overdosing, which can cause severe internal bleeding. According to Thummel, genetic testing could be cost-effective if it could replace some of the blood tests now used to monitor warfarin dosing—especially if it reduces hospitalizations by determining which patients need lower doses of the drug and closer monitoring. Other recent research may lead to improved dosing for the immunosuppressants cyclosporine and tacrolimus, which can cause kidney damage and failure.

Too much information? Knowledge of genetic susceptibility to smoking-related disease could actually decrease a person's motivation to quit.

Susceptibility to heart disease has also been linked to numerous genes, including *APOE*. Steve Humphries, a professor of cardiovascular genetics at University College London Medical School, reported that variants of *APOE* are well known to raise or lower blood levels of low-density lipoprotein ("bad") cholesterol, but only have a significant impact on an individual's risk of heart disease when the person smokes. Smoking increases the risk of heart disease of people with any of three common *APOE* variants, but the risk is greatest (about threefold) among carriers of the ε4 variant, which is found in about 25% of the population. Humphries and colleagues are working with a smoking cessation clinic in London to determine whether smokers will be motivated to quit if they learn that they have a high genetic risk of disease.

One possible complication is that information about *APOE4* status could lead to fatalism rather than a determination to quit smoking, said symposium co-organizer Wylie Burke, chair of the UW Department of Medical History and Ethics. The issue is further complicated by the fact that the *APOE4* polymorphism has also been linked to a higher risk of Alzheimer disease. People may view death by heart disease as a blessing compared to contracting Alzheimer disease, for which there is as yet no cure and no clear-cut means of prevention. Therefore, knowing one's own *APOE4* status could decrease, rather than increase, a person's motivation to quit smoking. In addition, according to Burke, focusing on genetic susceptibility to smoking-related disease may draw attention from more important environmental factors, such as advertising, that encourage people of all genotypes to begin smoking in the first place.

Many other findings in gene–environment interactions raise similar ethical, legal, and social issues about whether society should focus on labeling individuals as susceptible to a given disease or simply reduce environmental exposures for everyone. "We need to be very careful to create the right environment in this era of rapidly accumulating genetic information," said Burke. "An overemphasis on the effects of genes relative to the effects of the environment can lead to oversimplification of the problem and distract attention from needed environmental change." **–Kris Freeman**

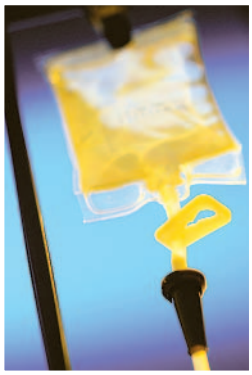
Clockwise from top left: Photodisc; Photodisc; Digital Vision; Photodisc

PROTEOMICS

The Bleeding Edge of Technology

In any given proteomics experiment, a cell or tissue can express hundreds or thousands of proteins at a time. Unlike the static genome, the proteome can change quickly, and key responses to toxicants and disease may involve small amounts of rare proteins. As an added complication, gene–protein and protein–protein interactions often are not linear. So what's a researcher to do? At the seminar "Proteomics and Systems Biology," held at the 2004 annual meeting of the American Association for the Advancement of Science in Seattle, Washington, presenters discussed advances at the "bleeding edge" of proteomics research and their use in the study of complex biochemical interactions within and among cells—advances that may help overcome some of the challenges posed by proteomics.

Seminar presenters discussed techniques to both measure protein signals and analyze the enormous amounts of data that result from such experiments. In most of the experiments reported, new techniques were tested on biological systems that had already been partially characterized, such as blood plasma. This allowed researchers to validate their systems with the bonus of potentially adding to knowledge about the biological systems in question.



In one analysis of blood plasma conducted at Pacific Northwest National Laboratory (PNNL), scientists claim to have identified about 3,700 proteins (not counting immunoglobulins) from human plasma, results that are an order of magnitude greater than those described only 18 months ago, according to Richard Smith, director of the NIH Proteomics Research Resource Center at PNNL. The newly detected proteins include many found at very low levels, some of which could be used as biomarkers of toxic exposure or disease progression, said Smith.

The plasma analysis used high-sensitivity, high-throughput instrumentation and techniques developed at PNNL, including Fourier transform ion cyclotron resonance (an advanced form of mass spectrometry). The PNNL researchers also separated out the most abundant proteins, allowing measurements to focus on less abundant proteins and increasing the number of proteins found in their plasma samples from about 1,000 to about 3,700. In addition, using electrospray ionization and low flow rates of solutions into the mass spectrometer facilitated detection of proteins in amounts as small as 10 zeptomoles, a level of sensitivity that makes it possible to analyze many proteins expressed by a single cell, said Smith.

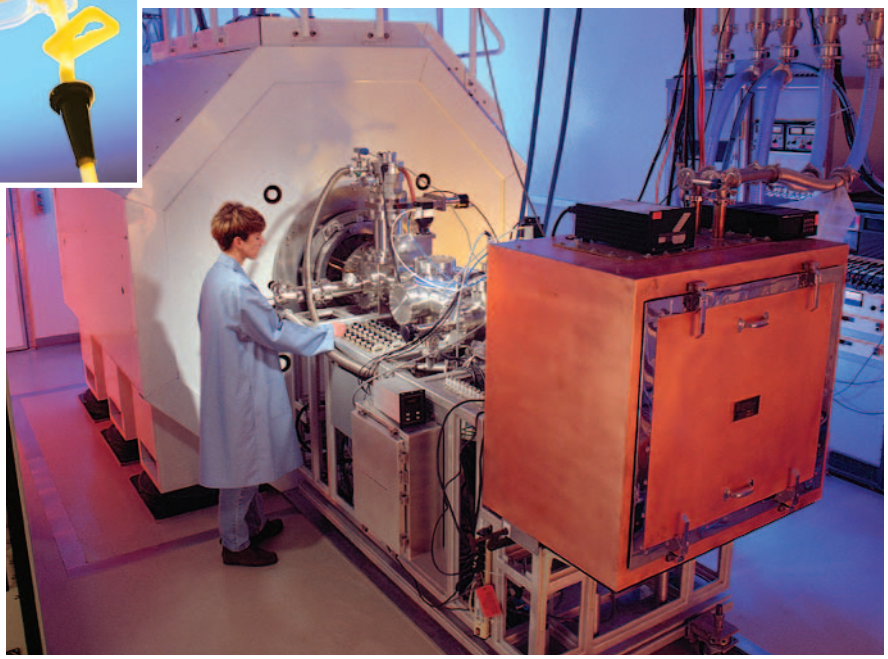
Leroy Hood, president of the nonprofit Institute for Systems Biology (ISB), too, discussed the potential for analyzing single cells. Researchers at the ISB and the California Institute of Technology are developing nanochips measuring 100 microns on a side that will assess the behavior of individual cells and gauge the concentrations of the mRNAs and proteins from a single cell. The ISB researchers have applied microfluidics—the study of how fluids behave at the nano level—to successfully conduct biological assays on single cells. ISB and Caltech researchers are now working on nanochips that can analyze several cells simultaneously. "We'll be able to interrogate a T cell and then an antigen-presenting cell separately. Then we will let the cells interact and interrogate their combined behaviors," said Hood.

Some presenters noted cautions about the prospects for analysis of single, living cells. Smith, for example, said that much more progress is needed in areas such as the construction of nanochips and microfluidics in order to make these methods of measurement "truly useful and not just a stunt."

Presenter Matthias Mann, a professor of biochemistry and molecular biology at the University of Southern Denmark, emphasized the need to distinguish between at least three states to track changes in protein expression. He displayed preliminary data from cells that were treated with a growth factor and sampled at five points. Analysis detected changes in levels of about 400 phosphoproteins over time. "The activation of some proteins decayed faster, and some proteins were activated later," Mann said.

Beyond the proteome is the metabolome—the sugars, amino acids, and other molecules that are created by or combine to create proteins. Masaru Tomita and colleagues at Keio University and the bioventure firm Human Metabolome Technologies are combining capillary electrophoresis and mass spectrometry to analyze the metabolites from rice, *Escherichia coli*, and *Bacillus subtilis*. For the purposes of these experiments, the team defined metabolites as molecules with a molecular weight of less than 1,000. In one experiment, the team detected more than 1,700 possible metabolites. The team is now constructing a model of entire metabolic pathways with several thousand reactions.

It may never be possible to fully describe the proteome of any species. Unlike the genome, which Hood characterized as "digital" and so "ultimately knowable," the proteome is affected by a multitude of external factors. In its absolute sense, according to Mann, the proteome is "as unreachable as the horizon." **–Kris Freeman**



Big picture, tiny molecules. A PNNL researcher performs an experiment on the Fourier transform ion cyclotron resonance mass spectrometer, which is used to characterize proteins such as those in blood plasma (inset).

PNNL; inset: Photodisc

NUTRIGENOMICS

Diet and DNA

The emerging field of nutrigenomics explores how nutrients in foods interact with genes that contribute to chronic diseases. The goal of nutrigenomics is to understand individual nutrient genotypes to design dietary interventions that restore health or prevent disease, eventually improving the health of the population at large as well as that of specific subpopulations. The fledgling field is packed with promise, and two new research initiatives aim to help deliver on that promise.

A Center of Excellence for Nutritional Genomics was established in 2003 at the University of California (UC), Davis, to coordinate nutrigenomics studies among participating institutes. A five-year, $6.5 million grant from the NIH National Center on Minority Health and Health Disparities funds the project. Genetics professor Raymond Rodriguez directs the new center, which unites 25 experts in nutrition, molecular biology, bioinformatics, and related fields from UC Davis, the Children's Hospital Oakland Research Institute, the U.S. Department of Agriculture Western Human Nutrition Research Center, and the Ethnic Health Institute at Alta Bates Summit Medical Center. Center members will explore how different foods interact with genes to increase the risk of type 2 diabetes mellitus, obesity, heart disease, and cancer.

Across the Atlantic, the European Nutrigenomics Organisation (NuGO) was launched in February 2004. This network of 22 scientists from 10 European countries will receive ¤17.3 million from the European Union over six years to develop new technologies, improve model systems, and advance nutritional bioinformatics. "Particular attention will be given to studies of human volunteers, and both biomarkers and new methods will be developed and validated," says Siân Astley, NuGO's communications manager.

"Nutritional genomics connects the Human Genome Project to human health in the most personal ways—through the foods we eat several times a day," says Rodriguez. "A better understanding of how diet and genes interact will enable us to better manage our own health and possibly prevent, mitigate, or delay the onset of chronic and age-related diseases."

People react to certain nutrients differently, depending on their genetic makeup. Lactose intolerance, a well-known example of nutrigenomics, afflicts largely Asians and Africans, and far fewer people of northern European descent. That's because a single base pair change in DNA occurred in northern Europe about 6,500–12,000 years ago, which allowed people there to digest lactose—in an environment with a short growing season, access to the additional nutritious food source of milk was helpful for survival, says Jim Kaput, a pioneer of nutrigenomics and founder of the diagnostics company NutraGenomics.

Today's nutrigenomics researchers hope to find gene variants that explain why, for instance, some people can lower their blood pressure through dietary changes, while others need drugs. Other variations might explain why some people are more susceptible to gastrointestinal cancers, inflammatory diseases, and osteoporosis.

Scientists are finding that biologically active components of foods can alter gene expression. For example, a deficiency of folic acid may lead to breaks in DNA that mimic radiation damage. Other nutrients are involved in molecular processes related to DNA structure, gene expression, and metabolism, which contribute to the development of chronic illnesses.

The nutrigenomics approach resembles pharmacogenomics, which looks at the relationship between single-nucleotide polymorphisms in genes and patients' responses to drugs to personalize medicine. Although progress in pharmacogenomics currently surpasses that in nutrigenomics, the two are closely linked. "Without nutrigenomics, pharmacogenomic data cannot be interpreted correctly, because diet may affect the expression of genes involved in drug metabolism," says Kaput. He proposes that pharmaceutical companies should include nutrigenomics in the design of new drugs because, he says, "what you eat affects a drug's efficacy." **–Carol Potera**

Lessons from lunch. The field of nutrigenomics is examining the nexus between diet and genetic susceptibility to disease.

Photodisc

RNAi

How Viruses Sabotage Silencing

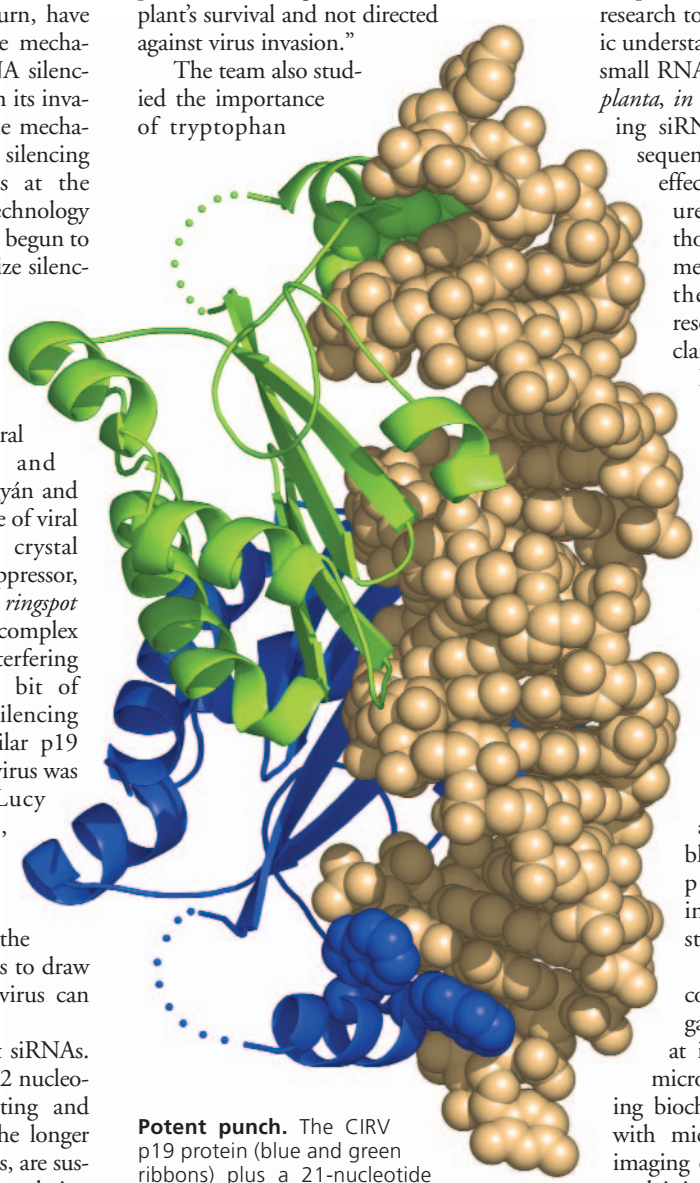
The discovery and description of RNA silencing less than a decade ago has spawned a flood of research, revolutionizing the practice of functional genomics and leading to intensive exploration of its potential application to treat numerous diseases. RNA silencing was first noticed in plants when attempts to create transgenic plants that overexpressed a natural gene often had the opposite effect; later it was found to be an evolutionarily conserved defense mechanism against plant RNA viruses and other molecular parasites. Viruses, in turn, have evolved their own counterdefense mechanisms: proteins that suppress RNA silencing, allowing the virus to maintain its invasion of a plant. Until recently, the mechanism behind this suppression of silencing was a mystery, but researchers at the NIEHS and the Agricultural Biotechnology Center in Gödöllő, Hungary, have begun to unravel how some viruses neutralize silencing, shedding important new light on a complex molecular interaction.

In the 26 December 2003 issue of *Cell*, NIEHS investigator Traci M. Tanaka Hall, postdoctoral researcher Jeffrey Vargason, and Hungarian colleagues József Burgyán and György Szittly elucidate the nature of viral counterdefense by solving the crystal structure of a known silencing suppressor, the tombusvirus *Carnation Italian ringspot virus* (CIRV) p19 protein, in complex with a 21-nucleotide small interfering RNA (siRNA), the workhorse bit of nucleic acid that drives the silencing process. The structure of a similar p19 protein found in another tombusvirus was published by Keqiong Ye, Lucy Malinina, and Dinshaw Patel, all of the Memorial Sloan-Kettering Cancer Center, in the 18/25 December 2003 issue of *Nature*. The slight differences in the structures have allowed researchers to draw further inferences about how a virus can interfere with RNA silencing.

There are two classes of plant siRNAs. The shorter ones, measuring 21–22 nucleotides, are responsible for detecting and destroying molecular invaders. The longer ones, measuring 24–26 nucleotides, are suspected to be more involved with regulating retrotransposons and DNA methylation. The structure of p19 reveals that the protein selectively recognizes silencing siRNAs by measuring their length. Tryptophan residues (Trp39 and Trp42) on the protein act like molecular calipers, forming a so-called stacking interaction with the ends of the end base pairs of the shorter siRNAs. By binding to the silencing siRNAs, the protein in effect sequesters them, rendering them incapable of carrying out their silencing mission and allowing the virus to run rampant within the plant.

The protein can also bind to the longer siRNAs, but much more weakly. "Viruses are smart," says structural biologist Tanaka Hall. "They don't want to kill their hosts too quickly. They want to have time to replicate. So it makes sense that the virus spares processes that might be essential to the plant's survival and not directed against virus invasion."

The team also studied the importance of tryptophan residues to p19's silencing suppression effects by introducing mutations into the p19 coding sequence of the CIRV genome to change one or both of the tryptophan residues. A protein database search had shown that both tryptophans were not absolutely conserved in all identified viral p19 sequences, although all sequences conserved at least one of the tryptophans, with the other residue usually capable of forming a stacking interaction. Their findings suggest that while the virus can at least partially succeed in suppressing silencing with two residues capable of stacking, substituting one with a glycine residue lacking a side chain results in failure of the suppression and recovery of the plant from infection.



Potent punch. The CIRV p19 protein (blue and green ribbons) plus a 21-nucleotide siRNA (beige balls) adds up to a lethal combination that enables viruses to counter a plant's attempt at RNA silencing.

Jeffrey Vargason

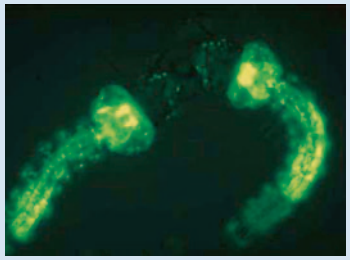
p19 has emerged as an important research tool in the quest to refine scientific understanding of how siRNAs and other small RNAs, such as microRNAs, work *in planta*, *in vitro*, and *in vivo*. Unlike silencing siRNAs themselves, which rely on sequence specificity to accomplish their effects, p19 appears to simply measure the siRNAs and neutralize those designed for silencing. This mechanistic revelation will enhance the value of the protein as researchers continue their efforts to clarify the biological roles of small RNAs in gene silencing and other cellular regulatory and developmental processes.

According to RNA silencing research pioneer Phillip Zamore, an associate professor in the Department of Biochemistry and Molecular Pharmacology at the University of Massachusetts Medical School, Tanaka Hall's structure of the p19 protein also provides great insight into the RNA silencing pathway itself. "All future models of the RNA interference pathway must incorporate a step at which they are vulnerable to siRNA sequestration by p19," he says. "This conclusion is inescapable once one has seen the structure."

Tanaka Hall's group plans to continue to use p19 in its investigations. "We're particularly looking at its role in being able to inhibit microRNA-initiated processes, looking biochemically at how p19 combines with microRNAs," she says. Structural imaging of that relationship could aid in explaining the still poorly understood mechanisms by which microRNAs accomplish their cellular tasks. **–Ernie Hood**

txgnet International Sequencing Consortium

The sequencing of multiple species' genomes by the Human Genome Project, including those of the human, the worm *Caenorhabditis elegans*, and the fruit fly *Drosophila melanogaster*, has laid the foundation for the field of comparative genomics. Experts believe this field of study represents the next step in genomic exploration, and that sequencing of more organisms will be critical to answering cross- and multispecies genomic questions. The International Sequencing Consortium (ISC) was established in 2002 to provide a worldwide forum for genomic sequencing groups and their funding agencies to share information, coordinate research efforts, and address common issues raised by genomic sequencing, such as data quality and release. The ISC has established a free website at **http://www.intlgenome.org/** where scientists and the public can get the latest information on the status of sequencing projects for the genomes of animals, plants, and other eukaryotic organisms.



The ISC website consists essentially of a database of sequencing projects. The database can be searched and sorted by organism, sequencing group, or funding agency. In addition, information is provided for each organism on the region of the genome being studied, strategies being used, the purpose of the study, collaborating groups, and the timetable for the project, if known.

In most cases, links in the database connect users directly to individual websites for the project, sequencing group, or funding agency. For example, the website lists a project for sequencing the mosquito species *Anopheles gambiae*, the principal vector of malaria. The project was conducted by Genoscope (the national center for sequencing in France) with funding by the U.S. Agency for International Development, the World Health Organization Special Programme for Research and Training in Tropical Diseases, and the French Ministry of Research. Another link on the ISC site connects to the completed project's website, where information is provided on the scope and importance of this work.

The ISC site also contains a listing of links to other resources categorized under the headings of microbial websites, genome browsers, trace archives, and other genome-related sites including other public databases where DNA sequence data are deposited.

Members of the ISC include large-scale, high-throughput sequencing centers and their funding agencies, all of whom have agreed to continue generating publicly available sequence data for unrestricted use by the research community. Most of the sequencing projects included in the ISC database adhere to the policy of rapid release of prepublication data that has been established by the National Human Genome Research Institute and The Wellcome Trust for efforts designated as "community resource projects." **–Kimberly G. Thigpen**

Top to bottom: Sinclair Stammers/WHO TDR; ISC; Sinclair Stammers/WHO TDR; A.W. Taylor-Robinson/WHO TDR