

Genome Technology

Inside Integrated Biology

THE BREAKTHROUGH CLUB



With two new CEGS centers, the breeding ground for next-generation genomics just got bigger. A look at these high-risk, high-reward propositions.

ALSO INSIDE:

Quantifying proteomics:
Label-free methods emerge

With Ohio program,
clusters go forth and conquer

Scientists speak:
Tips on sequencing cleanup

University of Washington's Maynard Olson and Deirdre Meldrum, directors of first-funded CEGS

**NEW:
SCIENTISTS
RECOMMEND
HOT PAPERS**

Building for Breakthroughs

Washington's Maynard Olson studies natural genetic variation



By Meredith W. Salisbury

A Democrat was president, Harold Varmus was director of NIH, and InforMax, reveling in its success, had just doubled its staff. It was the very late '90s, and it was becoming obvious even to skeptics that the human genome would in fact be fully sequenced in a matter of a year or two. For folks at the National Human Genome Research Institute, that meant the beginning of a new worry — what would come next?

“Clearly genomics was undergoing a big transition,” says Maynard Olson at the University of Washington. “With the Human Genome Project coming to fruition and very large-scale sequencing capacity having become routinely established, after ... years of there being an essentially monolithic goal for genomics, which was to sequence big genomes, ideas about where to go from there were diverse.”

“We were looking forward to what one might hope to achieve in future years once you have the genome,” says Jeff Schloss, program director at NHGRI. He and his colleagues there were seeking “promising ways for the institute to invest.”

With a plethora of medium-size genome centers scattered around the US, the answer seemed to lie in making use of those resources in a way that would encourage innovation and what Schloss hoped would be “the next generation of genomic technologies. Think about the impact that microarrays have had on biomedical research — ideally one would get something with that kind of impact out of this program.”

After plenty of discussion with the institute’s advisory council, that concept became known as CEGS, or the Centers of Excellence in Genomic Sciences. These centers, funded with as much as \$3 million per year for up to five years per grant — some co-funded by NIMH — were given a mandate to find a compelling and unique niche in integrative genomics. In return for the big bucks, NHGRI would accept riskier research pro-

U. Washington photos: Peter Yates

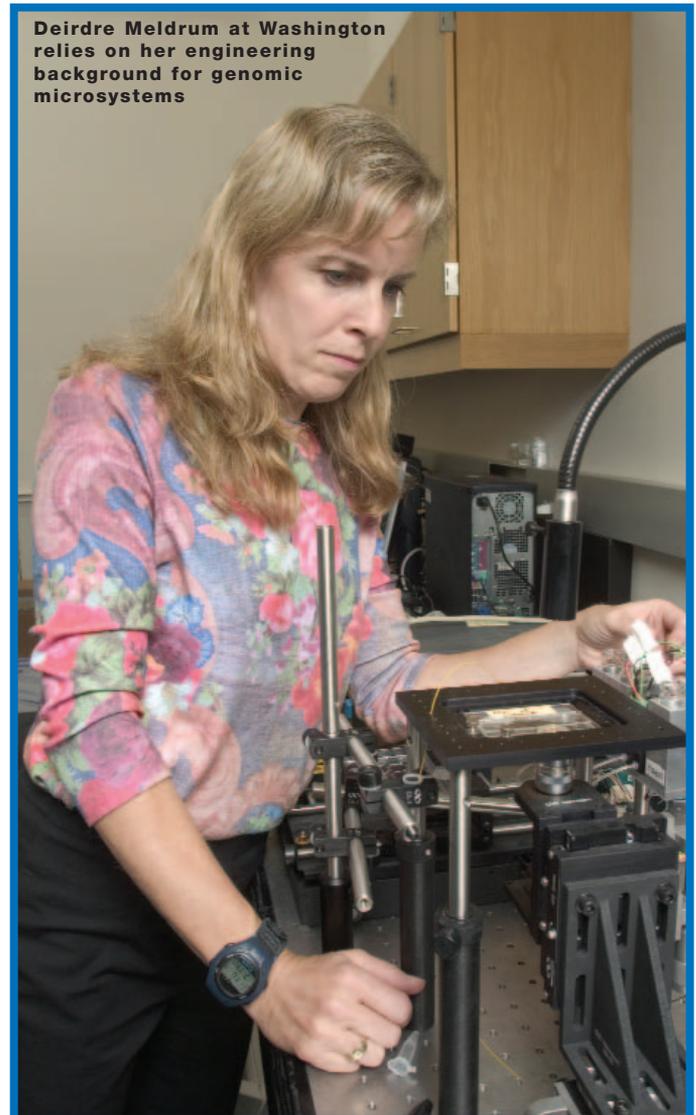
CEGS is a kind of Goldilocks program, building integrated biology centers that aren't too big and aren't too small. Though it's still too early to weigh in with a final verdict on the fruits of their science, the directors of these nine multimillion-dollar institutes are out to prove they're just right.

posals in hopes of leading to highly innovative technology and science. "They wanted 100 flowers to bloom to see what was successful," as Roger Brent, director of a CEGS based at the Molecular Sciences Institute, puts it. The first round of funding happened in late 2001, and with two new centers just announced this summer, the tally is up to nine CEGS operating out of universities and institutes across the country. (For more information on each center and the science going on there, check out the profiles on the following pages.)

"The program is a test, really, of the continuing relevance of what I would call medium-sized genome centers," Olson says. "The strength is it's a way of enabling moderately large-scale applications of genomics — not huge — in a big variety of areas, thereby to explore how effectively genomic approaches to a big variety of problems will play out."

"Variety" is certainly the right word. Because the centers' goals are far more flexible and investigator-initiated than many grants of this magnitude, the applications of genomics among them has diverged greatly. Of the nine CEGS, no two are alike; their research ranges from fish evolution to cell-measurement technology development to studying epigenetics in common diseases. "There's always been a need to preserve the democratic or pluralistic nature of the genomics community," says Richard Gibbs, who runs the Baylor sequencing center and chairs the review committee for the CEGS grants. NHGRI deliberately cast a broad net and has ponied up the funds to really back up these centers, although its commitment of \$25 million annually to the program still doesn't come near the funding it gives to production-scale sequencing centers and is not a major portion of its \$385 million budget this year for extramural research.

Still, as Olson points out, the diversity that he found so appealing about the CEGS program — his own center was funded in the first round in 2001 — might also be the program's greatest weakness: while each center is doing highly



Deirdre Meldrum at Washington relies on her engineering background for genomic microsystems

CEGSTech: Integrated Biologically Active Microsystems

Lead PI/Institution: Deirdre Meldrum, University of Washington

Grant term: Five years

Began: Summer 2001

Funds: \$15 million

Staff: 10 primary investigators, and around 40 postdocs, graduate students, and a few undergraduates

Goal: “To be able to measure multiple parameters in living cells in real time,” says Meldrum. Ultimately, she adds, that will lead to significant inroads into functional genomics. She also hopes the technology designed in her center will eventually be packed into microsystem modules that can be used by any researcher.

Scientific projects: Meldrum’s center is one of the most heavily focused on technology. A key component revolves around developing the technological microsystems needed to eventually enable asking biological questions. So far these systems include keeping a cell alive “under as natural conditions as possible in a microenvironmental chamber,” while performing single-cell proteomics and taking various measurements (such as cell oxygen respiration) and running RT-PCR in real time. Meldrum, an engineer by training, oversees the technology side

of things while co-director Mary Lidstrom heads up the biological applications. It’s a virtual center, Meldrum says, spread across 10 labs on UW’s campus and two at the Fred Hutchinson Cancer Research Center. Much of the tech will go toward biological questions aimed at using genomic and proteomic information to understand, for example, metabolic pathways, cell death, infection, and aging, Meldrum says.

Status: Some of the technology applications are “pretty far along,” Meldrum says. “The progress has just accelerated and we’re really making a lot of headway.”



CEGSTech co-directors
Deirdre Meldrum and Mary Lidstrom

valuable work, without any kind of programmatic goal, their results will not mesh nicely in any kind of grand finale.

The centers are marked by a number of traits that set them apart from many other genomics efforts: they’re all highly interdisciplinary, have a goal of fostering more collaborations than usual, and perform training and minority outreach as part of their mission.

PEOPLE WHO NEED PEOPLE

A major theme among these centers, spelled out in the program announcement, is that they must be “integrative” — drawing on people with varied backgrounds to bring genomics to bear on a number of new disciplines. “The program has been a good magnet for different kinds of talents that expand different aspects of genomics,” says Baylor’s Gibbs.

“We’re building a bridge with people from these greatly different disciplines,” says Brent at MSI. “The fact that we have real deep collaborative interactions between biologists and

people who have backgrounds in, say, experimental physics have enabled us to find out new things about the quantitative behavior of this particular signal transduction pathway. ... It’s even alerted us to new qualitative phenomena that we would not otherwise have been able to find.”

At the Johns Hopkins CEGS headed up by Andy Feinberg, there’s everyone from basic science geneticists and molecular biologists to clinical geneticists and genetic epidemiologists, says James Potash, director of disease applications there and a psychiatrist by training.

Of course, having the spectrum of backgrounds doesn’t mean much if the scientists can’t find ways to merge their work together. Deirdre Meldrum, who runs one of two CEGS based out of the University of Washington, says teamwork has been part of the vocabulary since day one. “That’s been a critical part, forming the team and finding people that like to work in a team,” she says. “People who really do spend the time and join in and really work together see how much you can gain [from that].” People at Meldrum’s CEGS are organ-

Molecular and Genomic Imaging Center

Lead PI/Institution: George Church, Harvard University

Grant term: Five years

Began: Summer 2004

Funds: \$10 million

Staff: A conglomeration of four labs, each with between five and 20 people



Goal: “The center’s focused on developing cost-effective technology for determining DNA sequence ... not just for the germline but also for cancer genomes and the immune system and so forth,” Church says.

Scientific projects: By now the genome community has become familiar with DNA sequence colonies, or polonies, the brainchild of Church and Rob Mitra. Still, Church says, few academic labs are working on what he believes promises to lead to a \$1,000 genome kind of technology. “We’re combining that [technology] with a biological direction which is also certainly not overrepresented in the genomics community, and that’s stem cells,” Church explains. He and fellow PIs including Mitra and David Gottlieb are collaborating both to achieve this low-cost sequencing tool and also to study RNA analysis, using polonies to do *in situ* analysis. They’ll use polonies to try to understand how stem cells differentiate, he adds. This center is heavily invested in technology development, something near and dear to Church’s heart. “My opinion since the beginning of the genome project is that the genome project would be well served having a good chunk of its money invested in technology development,” he says. “I’m attracted to the parts of the Human Genome Project that are aimed at increasing production rather than the producing.”

Status: While the polony research has been in the works for several years, this particular CEGS was actually funded just a number of weeks before press time.

ized into goal-oriented teams, and each team was designed to have an interdisciplinary mix of members, she adds.

That team mentality has also led to a number of collaborations, both between labs forming one CEGS and across CEGS. Brent’s center has a “coalition tentacle” at MIT, where researchers work on the heavily engineering side of things. At Yale, where Michael Snyder leads a CEGS with four other PIs, he says the collaborative nature among the five labs has “really brought together five people that don’t normally work together — now we meet many times a week. It really did bring us together to do things that we wouldn’t normally have done.”

RISKY SCIENCE

The complexity of research going on at any given CEGS is automatically higher just because there are scientists from so many disparate disciplines, says NHGRI’s Schloss. But beyond that, the CEGS are known for their riskier research proposals and a tendency toward innovation that most investigators can’t afford. First and foremost, the amount of money involved gives researchers “a chance to explore certain problems at a scale that would be hard to do in a regular R01,” Schloss says. With complexity and level of risk, he adds, “these are not by any means sure bets.”

At Stanford, that funding level meant a scientific team led by

William Talbot with co-PIs David Kingsley and Rick Myers could take the risky step of launching a new model organism, the three-spine stickleback. “That’s a risky, innovative project,” Talbot says. “To make a new model organism you’d need a number of individual grants, none of which would be all that compelling on its own.” The way he sees it, getting the CEGS grant enabled what might otherwise have been impossible with regular funding mechanisms.

Another appeal of the CEGS program was the length of the grant — five years, instead of the usual maximum of three. “That was another bold and attractive feature of this program,” Olson says. “Recognizing that these were going to be new entities, being built in most instances from scratch, and it was going to take a while for them to get off the ground and really be productive.”

REACHING OUT

With the additional resources offered by a CEGS grant come some additional stipulations, mainly in training and outreach requirements. Each center is responsible for including training programs and outreach programs to underrepresented minorities as part of the proposal, according to Schloss.

Though they all must address those requirements, the centers are free to handle them differently. At Brent’s CEGS, collaborators have started an ethics and policy course ➤ **p.34**

Center for the Epigenetics of Common Human Disease

Lead PI/Institution: Andrew Feinberg, Johns Hopkins University

Grant term: Five years

Began: Spring 2004

Funds: \$5 million

Goal: Building technologies and bioinformatic or statistical tools to better understand and characterize epigenetics, such as tools for allele-specific expression and the detection of methylation status, especially with regards to autism and bipolar disorder, says James Potash, director of disease applications at the center.

Scientific projects: “Epigenetics is a very hot area,” Potash says, defining it as “the study of information in a cell that is heritable during cell division but is not a feature of the DNA sequence itself.” That includes things like methylation, a type of natural gene silencing which has become better understood in recent years, and histone modifications, which have been studied less thoroughly, according to Potash. In order to allow for large-scale, high-throughput assessment of



epigenetic marks, he says, several new technologies are needed. “Now that the genome is completely sequenced, certainly one of the next frontiers is fully characterizing the epigenetic information across the genome,” he adds. Critical data on epigenetics comes from detecting and understanding methylation states and allele-specific expression — both of which primarily rely on technology that could use significant

ramping up, Potash notes. Lead PI Andy Feinberg and his team — including collaborators Matthew Portnoy, Gary Stormo at Washington University, and Webb Miller at Penn State — will also take a bioinformatics approach, trying to determine key sites or pockets of epigenetic activity by doing cross-species genome comparisons, Potash says. Additionally, “there are relatively few statistical tools developed to study epigenetic modification” — a problem Feinberg’s team hopes to remedy with statistical and modeling software to take into account epigenetic as well as other genetic information, such as SNPs. In the past, Potash says, epigenetics studies tended to focus on very rare diseases; Feinberg’s group will look instead at autism and bipolar disorder, two very common heritable diseases.

Status: Funding just started in April.

Genomic Approaches to Neuronal Diversity and Plasticity

Lead PI/Institution: Jingyue Ju, Columbia University

Grant term: Three years

Began: Fall 2003

Funds: \$11 million

Staff: 10 faculty-level members, with close to 50 postdocs, research scientists, and graduate students

Goal: “To discover the fundamental molecular mechanisms for long-term and short-term memory,” Ju says.

Scientific projects: Research at this center revolves around developing three new technologies, all linked to the underlying goal of better understanding memory. Working in conjunction with a scientific team at the University of Florida, Ju and company will design nanoscopic-level DNA arrays to use for global gene expression profiling; a chip-based massively parallel DNA sequencing system; and a

monitoring technology that can perform in real time to study mRNA species in live neurons. Ju uses *Aplysia*, a small organism with uniquely large cells, to test the technology. The tools will then be used to examine growth and plasticity in neurons and the molecular signals that interact with neurons.

Status: Despite having just a year of funding so far, Ju says that “all the projects are well underway.”



Implications of Haplotype Structure in the Human Genome

Lead PI/Institution: Michael Waterman, University of Southern California

Grant term: Five years

Began: Fall 2003

Funds: \$18.7 million

Goal: To better understand the haplotype structure in chromosomes and examine the structure's biological significance as well as its link to tracking down disease genes, according to NHGRI's summary of this CEGS.

Scientific projects: In this strongly multidisciplinary center, scientists are studying haplotype structure by designing algorithms to more accurately find polymorphisms through sequencing and chip-based research; looking into a link between haplotype blocks and recombination rates; and using population and association studies to determine possible explanations for haplotype evolution and to better approach studies of genetic traits, according to the NHGRI summary. (Waterman was unavailable for comment on his CEGS.)

age interaction among the CEGS members, rotating it around the sites to get more people involved.

Current CEGS leaders wax enthusiastic about their centers, and by all external accounts, deserve to be. All of the CEGS folks *Genome Technology* spoke with fully expect to send in proposals to renew their grants when the time comes. “By the end of next October we'll have submitted our competing renewal for another five years,” Meldrum says. “We would not be able to do this research without a big center grant.” **GT**

CENTER PROFILES CONTINUE THROUGH PAGE 36

focusing on genomics and citizenship at the University of California, Berkeley. The CEGS itself runs “a recruiting training program for underrepresented minorities,” Brent adds.

Yale's Snyder says his team offers lab sessions on how to use microarrays for undergraduates and students in the summer research programs — and even gives array demos for high school classes as well. At Stanford, Talbot and his colleagues teach a 10-day summertime course demonstrating lab exercises related to the stickleback and teaching techniques developed at the center. And in Washington, Meldrum says much of the training is internal — teaching engineers about biology and vice versa — but her center also trains incoming scientists. That program is partly fueled by a portion of the CEGS dedicated to providing seed funding that Meldrum and her colleagues can give out to new investigators.

FULL SPEED AHEAD

So far, CEGS leaders report that things are (finally) running like clockwork. But even as these centers continue to establish themselves, the exact future of the program remains uncertain. “The program announcement has formally expired at this point,” says Schloss. He and his colleagues at NHGRI are in discussions with their advisory council to figure out how to move forward. “There's little doubt that we'll continue the program, but we may change some things,” he adds. Rather than leave it as open-ended as it is now, “we may say there are some areas that need particular attention.” That discussion will flow into next year, since materials were not ready for a decision by the time talks wrapped up for grant season this year.

And in the meantime, CEGS leaders are working hard to build networks among themselves to make themselves more valuable to the community. Schloss says NHGRI runs a meeting for grantees each year to solicit feedback from and encour-

Genomic Basis of Vertebrate Diversity

Lead PI/Institution: William Talbot, Stanford University

Grant term: Five years

Began: Spring 2002

Funds: \$15 million

Staff: 20-25 people is a ballpark estimate; it's hard to quantify because some of the work has been done by subcontracting, Talbot says.

Goal: In a nutshell, "to understand how new vertebrate species evolve and how new vertebrate gene functions evolve," according to Talbot. In the process of doing that, the center is also "jump-starting a new model organism."

Scientific projects: Using two species of fish — zebrafish

and the three-spine stickleback — Talbot and the collaborating labs of Rick Myers and David Kingsley are studying evolution and gene function. The stickleback, Talbot says, is unique in its rapid and wildly varying evolution: members of the same species can look very different and have divergent behavior, but still successfully interbreed in a lab, Talbot adds. Though there's been tons of study of the fish from the ecological, evolutionary, and behavioral perspectives, "there had not been much molecular study," he says. Talbot and his colleagues are working furiously to build up the genetic resources for stickleback to get it up to the level of any other model organism, so they're sequencing ESTs, assembling BAC and cDNA libraries, building contig BAC maps, and cloning genes. Those efforts are largely to get the stickleback up to speed so that it can be studied as easily as

its distant cousin, zebrafish, which Talbot is hoping can be used in conjunction with stickleback to better understand duplicated genes and their outcome as well as to pick out genes responsible for rapid evolution.

Status: "Progress has been very good," Talbot reports. "A lot of the stickleback resources we set out to create are well underway and available in public databases." Going forward, the center is looking to continue building up these resources and use those to study the molecular basis of the evolutionary traits of interest.



Left to right: William Talbot, Rick Myers, David Kingsley

Center for Genomic Experimentation and Computation

Lead PI/Institution: Roger Brent, Molecular Sciences Institute

Grant term: Five years

Began: Summer 2002

Funds: \$15.5 million

Staff: Up to 40, including everyone; 20 who consider it their primary focus

Goal: To be able to predict a biological system — and get it right. "We [need] to figure out the trajectory down which cells go as well as the cells can figure it out," Brent says.

Scientific projects: "We can't really predict the future behavior of any well-studied biological system," says Brent, who spent years cobbling together this center, better known as

the "Alpha Project," from a variety of smaller grants until the CEGS grant came through. "We needed to be able to gain that ability." To that end, he and his team — as well as colleagues at Caltech, MIT, the University of California at Berkeley, and Pacific Northwest National Laboratory — are studying the pheromone response pathway in yeast to come up with new ways to predict biological systems. The idea is to come up with mechanistic models "which correctly predict the quantitative output of the system." In order to do that, he and his crew are coming up with new ways to measure cell activities such as protein movement or phosphorylation events. "The current high-throughput technologies cannot give you the data you need," he adds. "There is no chip, no mass spec that can do it."

Status: "We're just beginning to see the fruits now," he says. "There's a lot of important stuff that's going to come out this year."

Center for the Study of Natural Genetic Variation

Lead PI/Institution: Maynard Olson, University of Washington

Grant term: Five years

Began: Fall 2001

Funds: \$15 million

Staff: About 30 full-time, 45 total people involved

Goal: “You can learn a lot from looking at the DNA,” says Olson, “but ultimately you have to look at the people and assess phenotypes — connect this genotypic variation to phenotypic variation.” Olson’s group aims to better understand normal variation (as opposed to disease-based variation).

Scientific projects:

The technical focus, Olson says, is “on long-range resequencing of targeted segments of any genome.” Getting away from the relatively minimal base-pair limits of resequenced PCR products, Olson wants to be able to resequence hundreds of thousands or millions of base pairs in a human genome. Having that kind of population data enables his group to study the most variable sections of the human genome “such as microsatellites or triplet repeat polymorphisms,” Olson says.

“Are there pockets of very high SNP density in the human genome, and if so, do these account for the genetic variability that’s obvious if you look around at any small group of humans?” So far, these patches — they persist across tens of thousands of bases — of “hypervariability” have proven mysterious to Olson. They don’t appear to be under natural selection, he says, but there must be something that accounts for the high number of sections that have at least “10 times the characteristic levels of variation that you see everywhere in the human genome.” Olson adds, “We see it as part of a puzzle that we hope will lead to some clear insights.”

Status: “We’ve got all that technology working really quite routinely now,” Olson says. “We’re really at our peak operation. We have the basic techniques that we proposed to deploy really deployed and gathering data.”



Maynard Olson at Washington

Human Genome Array: Technology for Functional Analysis

Lead PI/Institution: Michael Snyder, Yale University

Grant term: Five years

Began: Fall 2001

Funds: \$15 million

Staff: Five PIs, and approximately 20 to 25 people all told, Snyder estimates.

Goal: After years spent developing technologies to map functional elements, in particular identifying genes and transcription factor binding sites, in yeast, Snyder and his colleagues are hoping to do the same for the human genome. Ultimately, they’d like to be able to “map the targets for any given transcription factor,” Snyder says.

Scientific projects:

“The theme of our grant is on using genome tiling arrays for finding new transcribed regions ... and for finding transcription factor binding sites,” Snyder explains. In the long run, he adds, the team — other PIs are Sherman Weissman, Rick Lifton, Mark Gerstein, and Perry Miller — is “trying to find, if you will, the goal of every single base of the human genome.” In first looking at human chromosome 22 using the genome tiling arrays, Snyder says,

“Much to our amazement, half the fragments that hybridized had not been annotated previously. ... We could show that there’s a lot more transcribed stuff out there.” Also, in mapping the first transcription factor for the chromosome they were studying, they saw binding “not only in the promoter proximal region but further upstream,” he says. Mutation mapping is another element of the technology development pipeline. In the process of all this, the team has worked on designing better microarray analysis tools that will prove universal for researchers in the field, and is now looking into a sample tracking database.

Status: Snyder’s team has had “a number of publications” so far from the work, and he says things are progressing nicely on getting the technologies established and working robustly.



Yale’s Michael Snyder