This month in Genome Technology

Also inside this month's Genome Technology

- **STRUCTURAL VARIATION**
  The latest findings and approaches to understanding copy number variation regions and what they mean.

- **METABOLOMICS**
  Work is underway to improve the technology and analytical tools needed to understand phenotype.
- **CASE STUDY**
  Researchers aim for molecular diagnostics to decipher the mystery of metastatic carcinomas of unknown primary.
- **CLINICAL Q&A: Gurvaneet Randhawa**
  The genomics advisor opines on evidence-based medicine.

Genome Technology Online

LOG ON NOW AT WWW.GENOME-TECHNOLOGY.COM

Explore all of our exclusive new resources for molecular biologists:

THE DAILY SCAN — Our editors' daily picks of what's worth reading on the Web

THE FORUM — Where scientists can inquire and comment on research, tools, and other topics

PODCASTS — Hear the full comments – in their own voices – of researchers interviewed in *Genome Technology* Magazine

POLLS — Make your opinions known on topics from the sublime to the ridiculous

ARCHIVE — Magazine subscribers can read every word of every article we have ever published

RECENT RESULTS FROM GT POLLS

How do you keep up with research in your field?

Response	%
I rely on PubMed. What else is there?	**68%**
I go online, to community-based sites like FriendFeed and blogs.	**10%**
I use referrals from other scientists.	**4%**
I stick to the basics: *Nature* and *Science*.	**18%**

1000-Series Thermal Cyclers

YIELD

SPEED

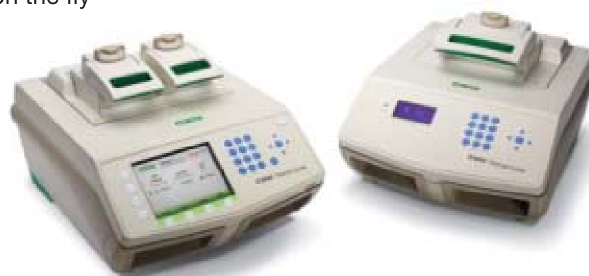
WE THINK YOU SHOULD BE ABLE TO OPTIMIZE ON THE FLY

Convenience Is What Happens When You Rethink PCR

Change the way you think about PCR with Bio-Rad's new family of thermal cyclers.

Wouldn't you rather optimize your reactions in minutes and not days? With Bio-Rad's new 1000-series thermal cyclers, optimizing on the fly is just the beginning.

- Easily interchangeable reaction modules meet any experimental or throughput need
- Reduced-mass sample blocks increase ramp rates and reduce run times
- Thermal gradient lets you incubate each row at a different temperature for fast protocol optimization



When you rethink PCR, you realize how easy it should be.

For more information, visit us on the Web at **www.bio-rad.com/pcr**

Purchase of this instrument conveys a limited non-transferable immunity from suit for the purchaser's own internal research and development and for use in applied fields other than Human In Vitro Diagnostics under one or more of U.S. Patents Nos. 5,656,493, 5,333,675, 5,475,610 (claims 1, 44, 158, 160–163 and 167 only), and 6,703,236 (claims 1–7 only), or corresponding claims in their non-U.S. counterparts, owned by Applera Corporation. No right is conveyed expressly, by implication or by estoppel under any other patent claim, such as claims to apparatus, reagents, kits, or methods such as 5' nuclease methods. Further information on purchasing licenses may be obtained by contacting the Director of Licensing, Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404, USA.

To find your local sales office, visit **www.bio-rad.com/contact/**
In the U.S., call toll free at 1-800-4BIORAD (1-800-424-6723)

Visit us on the Web at discover.bio-rad.com

BIO-RAD

BONUS: ARRAY CGH TECH GUIDE

Genome Technology

SEPTEMBER 2008

THE ART OF SPLICING

Scientists work to unravel alternative splicing

Copy Number Variation

Advances in Metabolomics

Informatics for Glycomics

DIANE LIPSCOMBE OF BROWN AND HARVARD'S MICHAEL WOLFE

Seize the Genome

NimbleGen Sequence Capture Arrays and Service

Maximize the power of next-generation sequencing by capturing and enriching specific regions of interest for targeted resequencing.

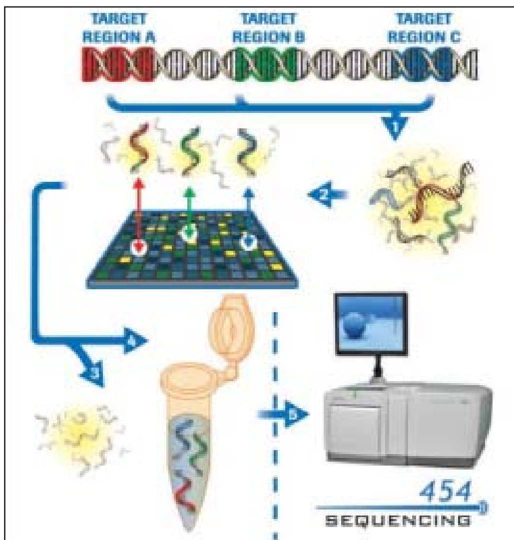


Figure 1: NimbleGen Sequence Capture Protocol
1. The genomic DNA sample is fragmented. 2. The sample is hybridized to a custom NimbleGen Sequence Capture array. 3. Unbound fragments are removed. 4. The target-enriched pool is eluted and amplified. 5. The enriched sample is ready for processing in the GS FLX sample processing workflow.

- **Target Specific Regions of Interest**
  Capture up to 5 Mb total sequence on a single array with high coverage and specificity.
- **Reduce Cost**
  Significantly reduce time and cost compared to laborious and limiting PCR-based methods.
- **Generate Data with Confidence**
  Ensure system performance prior to sequencing with built-in QC probes.
- **Customize Each Capture Design**
  Specify the array design to capture contiguous genomic regions or thousands of exons in parallel.

To seize command of your sequencing project,
visit **www.nimblegen.com/seqcap**
or call **(877) NimbleGen / (608) 218-7600**

454, 454 LIFE SCIENCES, and 454 SEQUENCING are trademarks of 454 Life Sciences Corporation, Branford, CT, USA, a Roche company. © 2008 Roche Diagnostics. All rights reserved.

Roche Diagnostics
Roche Applied Science
Indianapolis, Indiana

SEPTEMBER 2008

Contents

"We've long thought that the expression of different splice isoforms probably underlies a lot of the differential effects of drugs in different pathways."

Diane Lipscombe, Page 41

Metabolomics

Taming Metabolites

Metabolomics studies inch scientists ever closer to understanding phenotype. But to really make progress, pioneers are working on improving the technology and analytical tools of the field.

BY CIARA CURTIN

31

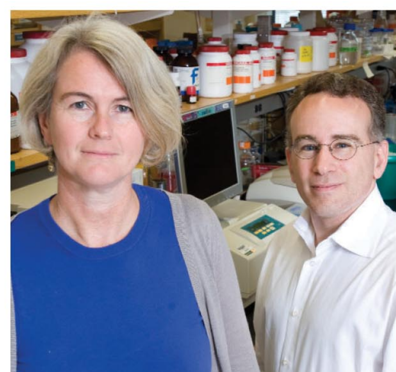
Structural variation

Counting on Copy Number

Research into copy number variation is highlighting just how complex genomic differences are. A glimpse at the latest findings and approaches to understanding CNV regions and what they mean.

BY MEREDITH SALISBURY

37



Alternative splicing

A Complexity that Goes Beyond Genes

Most genes are subject to alternative splicing, but it's still early days in understanding the phenomenon across pathways or on a genome-wide scale. A look at some pioneers in the field and the technologies they've commandeered to make sense of it.

BY JEANENE SWANSON

40

On The Cover

40 The Art of Splicing
37 Copy Number Variation
31 Advances in Metabolomics
25 Informatics for Glycomics

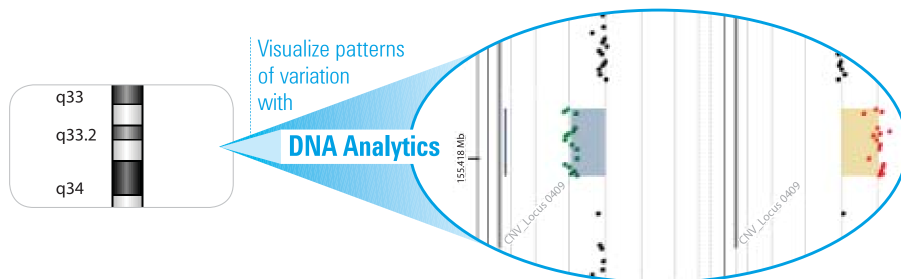
COVER AND TOP RIGHT PHOTOS: SAM RILEY

products | **applications** | software | services

OPEN TO:

Putting genomic variation in your sights.

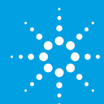
Growing numbers of scientists are turning to Copy Number Variants (CNVs) to gain a greater understanding of the links between genetic variation and disease. Agilent's proven CGH platform, combined with a newly expanded and validated probe database, enables you to interrogate CNV regions with greater sensitivity and specificity. DNA Analytics software provides you with a valuable new way to explore and visualize patterns of human genetic variation. Look into Agilent CNV microarrays—and see the whole picture on structural variation.



Learn what CNVs can mean to your research by visiting

www.opengenomics.com/cnv

© Agilent Technologies, Inc. 2008



Agilent Technologies

SEPTEMBER 2008

Contents

Markers

GWAS 10
NHGRI issues $31M to follow-up SNP research

SAMPLE PREP 11
Marziali's tech for DNA purification tackles the worst of contamination

GENE EXPRESSION 12
Cancer consortium traces signature for lung cancer recurrence

MICROFLUIDICS 13
RainDance readies to ship early-access version of droplet platform

SUPERCOMPUTING 16
New processor architecture holds promise for protein, gene studies

BOOK REVIEW 16
The Allen Reference Atlas: A Digital Color Brain Atlas of the C57Black/6J Male Mouse

CLINICAL LABS 17
AACC annual meeting covers genomics, proteomics, and diagnostics

In every issue

PRIMER 7
The simple life? Not for us

WHERE ARE THEY NOW? 9
ImmunoPCR, microarray diagnostics, and a whole new breed of institute

ZEITGEIST 18
Something to talk about

CAREERS 20
Savvy strategies in a budget crunch

INFORMATICS INSIDER 23
Homology: genealogy for genes

BRUTE FORCE 25
Sweet time for informatics

UNDER ONE ROOF 28
Starting from scratch

UPCOMING EVENTS 62

LAB REUNION 64
A man of many hats

BLUNT END 66

Upstream

SEQUENCING 49
Barcode of Life project leans on sequencing

PROTEOMICS 51
Satoris to launch Alzheimer's test

RNA INTERFERENCE 53
Tacere evaluates new drug opportunities

BIOINFORMATICS 55
Scientists debate publishing future

MICROARRAYS 57
For WTCCC, Agilent will provide custom chips

Downstream

PHARMACOGENOMICS 59
Pathwork first to net FDA OK for CUP Dx

CASE STUDY 60
Diagnosing the unknown

Q&A 61
Too much information

GENOME TECHNOLOGY PODCASTS:
A recent edition covered cloud computing, Esther Dyson, proteomics standards, and more.
www.genome-technology.com

DNASTAR®

Lasergene 8.0

BRIDGING THE GAP TO NEXT GENERATION SEQUENCE ANALYSIS

Meet the needs of today with a solution for tomorrow.

The Lasergene 8.0 suite of sequence analysis tools bridges the gap between conventional and next generation sequencing – allowing you to use the same software for analysis and visualization of both sequencing methods.

Featuring the highly effective functionality of previous versions, Lasergene 8.0 also includes many new capabilities.

New Lasergene 8.0 benefits your research with:

- Simplified primer design and virtual cloning, thanks to the added primer designing capabilities in SeqBuilder
- Faster and easier SNP detection and analysis for your specific project needs
- Ideal for Next Generation platforms; can be used with SeqMan NGen for enhanced analysis and visualization of data

Discover how Lasergene 8.0 is bridging the gap between the needs of today and the promise of tomorrow. Contact us for a free 30-day fully functional trial version.

DNASTAR, Inc • Madison, WI, USA • www.dnastar.com
Toll free: 866.511.5090 • 608.258.7420 • UK 0.808.234.1643

Genome Technology
ISSUE NO. 84

125 Maiden Lane, Second Floor
New York, NY 10038
Tel +1 212 269 4747 Fax +1 212 269 3686
GENOME-TECHNOLOGY.COM

EDITORIAL DIRECTOR
Meredith W. Salisbury
msalisbury@genomeweb.com

SENIOR EDITOR
Ciara Curtin ccurtin@genomeweb.com

SENIOR WRITER
Jeanene Swanson jswanson@genomeweb.com

REPORTER
Matthew Dublin mdublin@genomeweb.com

ART DIRECTOR
Paul J. Schindel pschindel@threebears.com

DESIGN & PRODUCTION
Lois Savander • John Burrows

GENOMEWEB DAILY NEWS
Bernadette Toner, Editorial Director, News
btoner@genomeweb.com
Ed Winnick, Managing Editor
ewinnick@genomeweb.com
Matt Jones, Reporter
mjjones@genomeweb.com
Andrea Anderson, Reporter
anderson@genomeweb.com

GENOMEWEB NEWSLETTERS
Kirell Lakhman, News Editor
klakhman@genomeweb.com
Justin Petrone, BioArray News Editor
jpetrone@genomeweb.com
Vivien Marx, BioInform Editor
vmarx@genomeweb.com
Alex Philippidis, BioRegion News Editor
aphilippidis@genomeweb.com
Ben Butkus, Biotech Transfer Week Editor
bbutkus@genomeweb.com
Charlotte LoBuono, Cell-Based Assay News Editor
clobuono@genomeweb.com
Julia Karow, In Sequence Editor
jkarow@genomeweb.com
Turna Ray, Pharmacogenomics Reporter Editor
tray@genomeweb.com
Tony Fong, ProteoMonitor Editor
tfong@genomeweb.com
Doug Macron, RNAi News Editor
dmacron@genomeweb.com

CHAIRMAN AND PUBLISHER
Dennis P. Waters, PhD
dwaters@genomeweb.com

ADVERTISING/SALES
Judy Block, Associate Publisher
jblock@genomeweb.com +1 212 651 5629
Allan Nixon, Director of Business Development
anixon@genomeweb.com +1 212 651 5623
David Samuels, Subscriptions and
Site License Manager
dsamuels@genomeweb.com

OPERATIONS AND FINANCE
Greg Anderson, Director
ganderson@genomeweb.com
Philip Borowiecki, Associate
pborowiecki@genomeweb.com

Copyright ©2008 GenomeWeb LLC.
THE GENOMEWEB INTELLIGENCE NETWORK
No portion of this publication may be reproduced in whole or in part without written permission from GenomeWeb.

The Simple Life? Not for Us

In looking over the pages of this issue of *Genome Technology*, it occurs to me that there's a theme here — however unintentional it may have been when we started. Many of the articles this month revolve around the concept of complexity.

Remember all those years ago when Incyte boasted about having 100,000 human genes in its database? It was quite a disappointment when the community homed in on the real number of human genes, leaving our egos just a bit bruised as we wondered how we could be such delightfully complex creatures when we had the same gene count as a mouse (or, for that matter, *Arabidopsis*).

As *GT* readers have known for a long time, there's a lot more than gene count factoring into genetic diversity. Our cover story delves into the world of alternative splicing, a genomic phenomenon that allows us to be economical in number of genes but without skimping on the products they encode. As Jeanene Swanson reports, scientists are using a range of technologies to study why and how alternate splicing takes place — as well as the effect it has on organism development. Research in this field has led to a better understanding of diseases, particularly in neurodegenerative conditions that have proven difficult to make sense of with other approaches.

While splicing adds a significant layer of complexity to genomic studies, so too does copy number variation, the focus of a feature story in this issue. While we may have just 20,000 (ish) genes, we're sneaky with them, packing our genomes with copies of the same genes. Sometimes they're inverted or changed ever so slightly, as if our genome was afraid of getting caught stacking the deck. In our article on CNV, we checked in with scientists leading the field to find out more about increasing use of variation studies in model organisms, the tools needed to accurately and comprehensively find gene copies, and how the mechanism has helped scientists establish links to disease. One expert we interviewed, Harvard's Charles Lee, went so far as to predict that a clear link of cancer predisposition with copy number variation is "right around the corner."

Of course, complexity goes way beyond the genomic level. Ciara Curtin reports on the growing field of metabolomics. Her feature story describes the advances researchers have made in the detection and deconvolution of metabolites, while noting that they're still grappling with establishing more comprehensive databases. In his Brute Force column, Matt Dublin considers the nascent glycomics space, focusing on the informatics at play there as well as on the community's attempt to implement standards early in the game.

Meredith W. Salisbury, Editor

Research Development Manufacturing

The Genomics Services Company

Cogenics is setting the standard in customizing and delivering expert genomics solutions for Research, Clinical, and Manufacturing applications in the biotechnology and pharmaceutical industries.

Whether your questions are best answered by sequencing, conventional or next-generation, gene expression, genotyping, or a combination of techniques, Cogenics provides resource-effective, expertly-run solutions for your research or FDA regulated genomics projects.

Your analyses will be performed using the most appropriate platform to answer your research questions with fast delivery times and high quality data. Whether you are planning a full or pilot project, here are some of the solutions we consistently provide:

- Sequencing solutions
- Viral and oncogene analyses
- SNP Discovery and Genotyping
- Cell Bank Characterization
- Genetic variant assay development and validation
- Drug efficacy and safety related analyses
- Support of global multi-center clinical trials
- Biodistribution and Residual DNA Analyses

www.cogenics.com/gogreen

US: 1 877-226-4364
UK: +44 (0) 1279-873837
Email: sales@cogenics.com

France: +33 (0) 456-381102
Germany: +49 (0) 8158-998518
www.cogenics.com

COGENICS™
The Genomics Services Company

>GT ONLINE
www.genome-technology.com

GT ONLINE POLL RESULTS:
Grants are tougher to come by; what's the best approach to still get one?

17%
SUBMIT AT LEAST TRIPLE THE USUAL NUMBER OF GRANT APPLICATIONS

2%
ASK FOR LESS MONEY AND YOU'RE MORE LIKELY TO GET FUNDED

43%
IT'S NOT ABOUT MORE GRANT APPS, BUT MORE TARGETED ONES; SPEND MORE TIME CRAFTING YOUR APPLICATIONS

37%
PARTNER WITH SOMEONE ELSE AND SUBMIT A JOINT PROPOSAL

>NEED HELP?

CHANGE OF ADDRESS — To change your address or manage your subscription to *Genome Technology*, the Daily Scan, or any Genome-Web publication, sign in to www.genome-technology.com and click on My Account, or e-mail us at evolving@genomeweb.com.

FREE SUBSCRIPTION — *Genome Technology* is free to active researchers in the life sciences. To subscribe, log on to www.genome-technology.com and fill out a 30-second form.

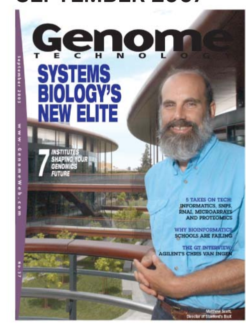
REPRINTS — For permission to reproduce material from *Genome Technology*, or to order offset-printed or Web-ready PDF reprints, write to reprints@genomeweb.com or call +1.212.651.5632.

ARCHIVE ACCESS — Subscribers can search through the complete *GT* archives by issue or keyword. Simply log in on the website.

ImmunoPCR, Microarray Diagnostics, and a Whole New Breed of Institute

A year ago in *Genome Technology*, a feature story checked into immunoPCR as an up-and-coming diagnostic tool. Only 15 years old, the technique couples an antibody-based detection assay with real-time PCR, allowing for the detection of proteins and viruses in very small concentrations. At the time, TATAA's Mikael Kubista was tweaking his PSA test to use immunoPCR instead of ELISA, and just one company — Chimera Biotec — was selling customized immunoPCR kits. While the technique continues to make headway in the lab — it's used in drug discovery, biomarker detection, and as a clinical lab tool — tricky conjugation protocols have made it difficult to standardize and move beyond the bench.

SEPTEMBER 2007



SEPTEMBER 2003

Last year's cover story took an eagle's eye view of the microarray diagnostics arena, reviewing the main players and the chip-based assays they were developing. Roche's AmpliChip CYP450 and Agendia's MammaPrint were the first array-based diagnostic assays to gain FDA approval. Others were hard at work creating their own tests: Heidi Rehm at Harvard-Partners Center for Genetics and Genomics was developing tests for hearing loss and for hypertrophic cardiomyopathy, and Emory's Madhuri Hegde was developing one for X-linked muscular dystrophy. In July of this year, Pathwork Diagnostics' Tissue of Origin test became the second *in vitro* diagnostic multivariate index assay to be cleared by the FDA.

In September 2003, our cover story delved into the rise of the systems biology community by tracking seven pack leaders. The story looked at the leadership, structure, and goals of these new interdisciplinary biology centers, which included Stanford's Bio-X Program for Bioengineering, Biomedicine, and Biosciences; MIT's Computational and Systems Biology Initiative; Princeton's Lewis-Sigler Institute for Integrative Genomics; Duke's Institute for Genome Sciences and Policy; the University of Michigan's Life Sciences Institute; the California Institute for Quantitative Biomedical Research; and the Cornell Life Sciences Initiative. These have heralded a wave of similar systems biology institutes focused on meshing different disciplines for large-scale, collaborative research. (The trend has become so popular that *GT* has since created a special column that profiles a new systems biology center each month.) Some of the multimillion dollar institutes that have graced our pages this year include Barcelona's Centre for Genomic Regulation, the Michael Smith Genome Sciences Centre in Canada, and the Burnham Institute in California.

— *Jeanene Swanson*

Markers NEWS

>SHORT READS

Last month, Alan Guttmacher took on his new role as acting director of NHGRI after Francis Collins officially stepped down. Guttmacher is a pediatrician and medical geneticist, and has been deputy director of the genome institute for more than five years.

The US Departments of Energy and Agriculture will give nearly $11 million over three years to fund 10 genomics research programs that can help develop bio-energy feedstocks for use in cellulosic biofuels. Recipients include the University of Georgia, Penn State, Michigan State, Oregon State, and the University of Massachusetts, among others.

In response to the growing number of so-called "minimum information" checklists for high-throughput biology experiments, an effort known as Minimum Information about a Biomedical or Biological Investigation is attempting to provide a set of guidelines for standardization groups to ensure inter-operability and prevent duplication of effort.

Affymetrix acquired True Materials, a firm with microparticle technology for use in diagnostic applications, for about $25 million in cash. Randy True, founder of the San Francisco startup, joined Affy as vice president of research and development.

GWAS: NHGRI Issues $31M to Follow-Up SNP Research

Flush with $31 million in funding from the National Human Genome Research Institute, four researchers will be following up on putative disease-related SNPs identified through genome-wide association studies. These scientists will be examining the prevalence of common disease SNPs in different, already existing epidemiological cohorts to see if they are still linked to disease in general populations. "If you think about a typical genome-wide association study as looking at many genetic variants in relation to one or a few health outcomes, in this follow-up program, we're looking at one or a few genetic variants in relation to many health-related outcomes," says Lucia Hindorff, program director at NHGRI.

GWAS research uncovers genes associated with disease, but does so in a population predisposed to find such SNPs, even in case-controlled studies. "Which is great for discovery," says Vanderbilt University's Dana Crawford, who received $7 million of this grant. "But when you're trying to describe what the SNP looks like in the general population — we don't have any data on that yet."

> **"We hope to be able to fill in gaps about what's known."**

The cohort that Crawford will be working with is from a long-running Centers for Disease Control and Prevention cross-sectional study that, as she says, takes a slice of America. Her population is one-third each Mexican-American, African-American, and European-American. Most of the GWAS SNP discoveries so far have been conducted in populations of European descent, and looking in her cohort could show if the SNP-disease association holds true in a more diverse population.

DANA CRAWFORD

Not all cohorts under investigation are cross-sectional; others have followed participants for decades. Fred Hutchinson Cancer Research Center's Charles Kooperberg will be studying SNPs in the Women's Health Initiative population, which has followed 160,000 women since 1991. "We have a wealth of epidemiological data. We have loads of clinical outcomes data," he says.

Just which SNPs and what epidemiological data will be the focus, Crawford and Kooperberg don't yet know, but the common diseases are likely to include diabetes, heart disease, and cancer. Crawford and Kooperberg, along with the Gerardo Heiss at the University of North Carolina, Chapel Hill, and Loïc le Marchand at the University of Hawaii Research Cancer Center will act as a consortium and soon will meet to prioritize a list of SNPs. "We hope to be able to fill in gaps about what's known about genetic variants in these populations with this program," says Hindorff. *— Ciara Curtin*

NEWS | Markers

Sample Prep: Marziali's Tech For DNA Purification Tackles The Worst of Contamination

It's not every day you see biologists get really excited about — of all things — sample prep. But if Andre Marziali has any say in it, you can expect that kind of enthusiasm to become downright commonplace.

Marziali is the brain behind SCODA, a nucleic acid separation technology that relies on special electrical fields to purify DNA or RNA out of even highly contaminated samples. The technology has just been exclusively licensed from Marziali's research home at the University of British Columbia to his startup, Boreal Genomics, for commercialization.

By the start of this year, Marziali already had prototype instruments out in the field, where researchers were getting acquainted with the gel-based method that uses focusing electrical fields, which nucleic acids respond to but contaminants don't. The fields concentrate the DNA or RNA into a compact unit at the center of the gel, while all other molecules are spun out to the periphery of the gel. In a metagenomics project with Rob Holt, Marziali describes taking a soil sample that no one else had been able to extract DNA from. Using the SCODA technology, Marziali's team was able to deliver a purified microgram of DNA — enough to get the project going.

ANDRE MARZIALI

Currently, Marziali is working on a second-generation prototype, which would speed up the purification process from a few hours to as little as five minutes. Boreal is hoping to get a beta version of that instrument out to customers by early next year. The cost will be comparable to similar electophoretic products, Marziali says: "In the future we imagine an instrument priced below $10,000."

The tool, and the concept behind it, has proven so popular that Marziali finds himself in the unusual situation of having to "turn investors away." Tom Willis, who chairs Boreal's board of directors, helped Marziali get the company off the ground and work out a business model relying on as little venture capital as possible.

> **"When we first started applying this to DNA, we didn't know how selective it was going to be. It's just played out beautifully."**

The technology, which can take any number of sample types from soil to blood to milk, is broadly applicable in forensics, agriculture, genomics, and throughout the life sciences.

Marziali says he stumbled across the idea for SCODA quite by accident. "When we first started applying this to DNA," he says, "we didn't know how selective it was going to be. It's just played out beautifully." — *Meredith Salisbury*

> SHORT READS

Illumina completed its acquisition of Avantome, a privately held developer of low-cost, long read-length sequencing technology, for $25 million in up-front payments and contingent payments of as much as $35 million. The startup was co-founded by Stanford's Mostafa Ronaghi, who joins Illumina as senior vice president and chief technical officer. Separately, Illumina announced that it named Stephen Pentoney to the post of VP for assay and reagent development in life sciences.

Joanne Sun is now director of protein analytics and high-throughput purification at the antibody discovery company Adimab. She formerly worked in clinical trials for Adnexus Therapeutics and in pre-clinical development at Abbott.

A group of scientists led by NHGRI's Colleen McBride published a commentary in *Nature Genetics* calling for an increased emphasis on translational research to make sure that personalized genomics delivers on its promise. They contend that regulators and members of the biomedical communities need a better understanding of how information from genetic tests is used in the clinic, how useful those tests are, and how genetic knowledge is viewed and used by patients before genetic technologies find their way into mainstream use.

Markers | NEWS

>SHORT READS

An international team of researchers from Germany, the US, Croatia, and Finland used the Roche 454 sequencing platform to sequence the Neanderthal mitochondrial genome to about 35 times coverage. The publication in *Cell* was based on DNA isolated from a bone more than 38,000 years old that was discovered in Vindija Cave in Croatia in 1980.

The Washington University School of Medicine in St. Louis hired Barry Sleckman as director of the Division of Genomic Medicine. Sleckman, who joined the school 10 years ago as an assistant professor of pathology and immunology, studies DNA repair and development of the early immune system.

BioNanomatrix received a $399,020 grant from NHGRI to continue development of its nanoscale imaging platform for haplotyping and gene mapping in a massively parallel format. This is the fourth grant the company has won from NIH to work on the technology.

The Broad Institute of MIT and Harvard released its Integrative Genomics Viewer, an informatics tool that allows scientists to visualize genomic data. The tool's zooming and panning abilities make it feel like Google Maps, according to Broad scientists.

Gene Expression: Cancer Consortium Traces Signature For Lung Cancer Recurrence

Early-stage lung cancer patients are usually treated surgically and sent home, but 25 percent to 30 percent of those people will have their cancer recur. Researchers from cancer institutes in the United States and Canada are paving the way to develop gene expression panels to predict which lung cancer patients will relapse. "If you know who these high-risk patients are, then you would potentially be able to offer them additional therapy," says David Beer, a professor of surgery and radiation oncology at the University of Michigan.

To develop a gene expression panel that can predict which lung cancer patients have a poorer prognosis than others, researchers need a lot of samples to analyze for markers — more than they typically have access to. To manage that, Michigan, H. Lee Moffitt, Memorial Sloan-Kettering, and Dana-Farber formed a consortium to track down a gene signature for lung cancer prognosis. "We really need to look at a larger number of tumors," says Beer. "We decided to basically pool our resources and put all of our tumors together."

"We really need to look at a larger number of tumors."

In this three-year study, the researchers collected 442 lung cancer samples from different treatment sites. Each location used the same conditions for isolating samples, the same criteria for inclusion, and reagents from the same production lot. Samples from two of those sites became the training set, and researchers from the four institutions independently used those samples to build predictive models. "The models that we tested were very, very broad from one gene to thousands of genes, and they all encompassed different approaches," Beer says.

DAVID BEER

Of the four panels developed, two were able to predict prognosis in the independent, blinded test sets. Furthermore, when the gene expression panels were used in conjunction with the clinical data, they worked even better. "That meant that putting the two together is more powerful than using just one alone," Beer says. The results are reported in the August issue of *Nature Medicine*.

While not yet ready for the clinic, these results indicate that "there is information in the genes that can tell you about the behavior of tumors," says Beer. He and his colleagues are now working to determine whether the morphological heterogeneity of lung cancer tumors has a molecular basis. "It's possible there's a lot of different ways that lung cancer becomes a cancer," he says. "Genomically it's a more difficult problem to unravel that."

— Ciara Curtin

NEWS Markers

Microfluidics: RainDance Readies to Ship Early-Access Version of Droplet Platform

If all goes according to plan, RainDance Technologies aims to get its droplet-based microfluidics tool into the hands of early-access researchers this fall, says Steve Becker, vice president of commercial operations.

Becker says that the main challenges in high-throughput biology today are miniaturization, automation, and multiplexing — all issues that he believes the company's RainStorm technology addresses head-on. "Doing research in droplets allows people to go back to what I'll call simplicity," he says. "Each droplet is the functional equivalent of a test tube or a well." The droplets can be processed at speeds of 3,000 per second, or more than 10 million samples in an hour. "Having that kind of throughput allows you to do ultimately single-cell or single-molecule [experiments]," he adds.

The technology will first be launched for the targeted resequencing market. Becker says that while next-generation sequencing is "growing at an unprecedented rate," scientists have not yet had an efficient way to perform genome enrichment for resequencing. Using RainDance's tool, scientists would have one primer pair per droplet for as many regions as desired, and then they'd perform "good old-fashioned PCR" in emulsions within those droplets. Becker contends that recent approaches of doing this with arrays or in solution "create bias" but that "using a very well-referenced PCR" would lead to a far less biased product.

STEVE BECKER

They'll work to show evidence of that in a partnership with Scripps, where scientists will use the RainStorm technology for targeted resequencing.

The droplets can be used to perform "just about every general lab application," Becker says; they're thermostable and biocompatible, as well as consistent at the picoliter size. "We're able to pack many of these droplets next to each other and they will not coalesce" — unless, that is, you want them to. The droplets can be merged on demand, or sorted "using [a] soluble fluorescent protein marker and a laser," Becker adds.

RainDance was founded in 2004, in part by 454 Life Sciences' Jonathan Rothberg, who is now chairman of the board. The company is currently developing and manufacturing the droplet platform that will go out to customers. According to Becker, scientists will send RainDance their list of loci of interest; the company will create a library of droplets with the corresponding primers; and then the library will be shipped to the customers for use in their labs.

— *Meredith Salisbury*

> **"Doing research in droplets allows people to go back to what I'll call simplicity."**

> SHORT READS

Celera named Jean Amos Wilson as president of laboratory operations at its Berkeley HeartLab subsidiary. She was previously senior director of genetic services at Sequenom.

Researchers from the University of Toronto used high-density oligonucleotide arrays to look at CNV frequency in the general population and in families with Li-Fraumeni, a syndrome predisposing individuals to cancer, finding that people with more CNVs in their genomes may also be at increased risk for cancer. The group, led by senior author David Malkin, published in the *Proceedings of the National Academy of Sciences*.

Larry Wellman joins OpGen as HR veep. He was previously in the same position at Digene, now part of Qiagen.

BioTrove hired Derek Potter as director of European business operations. Potter was previously European sales manager for Applied Biosystems, and he was involved in establishing Fluidigm's European operations.

OpGen, a single-molecule analysis company based in Madison, Wis., will open a facility in Gaithersburg, Md. The company says it expects to hire 80 people in the next two years to staff the facility, which will house R&D, service operations, and manufacturing.

10. What do you like best about EA?
QUICK TURN AROUND AND HIGH QUALITY WORK

Scientist, Top 10 Pharmaceutical Company

10. What do you like best about EA?
Good communications
Reliable processing

Associate Director, Clinical Programs, Pathwork Diagnostics

Comments:
Great job on the RNA isolations, the RNA looks great!

Senior Research Associate, Bay Area Biotechnology Company

6. How likely are you to recommend EA to a colleague? Would you say the chances are...
Excellent Very Good Good Fair Poor
I already did (2x)

Associate Director, Cambridge, MA Pharma Company

9. What additional services or platforms would you like EA to offer?
We are very satisfied with the range of services & platforms.

10. What do you like best about EA?
The responsiveness of the employees and the confidence in the quality control.

11. In what areas can EA improve?
None known.

Executive Vice President, Operations
International Neuroscience Network Foundation

Comments:
I like my relationship with EA where I feel comfortable calling Steve or anyone else to discuss topics. Open communication is good.
Tx.

Steve Tirrell, Millennium Pharmaceuticals, Inc.

EXPRESSIONS *of*

Expression Analysis Satisfaction Survey — Summer 2008

What your peers are saying about Expression Analysis is the very thing *you* could be saying. Bring us any project, and experience EA's superior quality for yourself. You too will enjoy a rewarding feeling of total confidence.

www.ExpressionAnalysis.com/CONFIDENCE

Expression Profiling • Genotyping • GWAS • Bioinformatics • Candidate Gene Resequencing • Next Gen Sequencing

9. What additional services or platforms would you like EA to offer?

I am satisfied with EA very much.

10. What do you like best about EA?

Before/After care about their client.

Research Fellow, NIDCR

10. What do you like best about EA?

QUICK TURNAROUND TIME
VERY FLEXIBLE
EXCELLENT PRICING

Scientist, Top 10 Pharmaceutical Company

10. What do you like best about EA?

Quality of work, turn around time, quick response to questions, excellent communication.

Senior Scientist, IPSEN Biomeasure, Inc.

12. On a scale of 1- 5 where 1 represents "Extremely Dissatisfied" and 5 represents "Extremely Satisfied" please rate the following:

a. Responsiveness to your needs	1	2	3	4	5
b. Degree of focus on your objectives	1	2	3	4	5
c. Openness and level of communication	1	2	3	4	5
d. Level of expertise	1	2	3	4	5
e. Degree of collaboration with you/staff	1	2	3	4	5
f. Quality processes	1	2	3	4	5
g. Ability to appropriately resolve issues	1	2	3	4	5
h. Adherence to deadlines	1	2	3	4	5
i. Quality of final deliverables	1	2	3	4	5
j. Delivery of actionable results	1	2	3	4	5
k. Quality of customer service	1	2	3	4	5
l. Competitive pricing	1	2	3	4	5

Scientist, RTP Pharmaceutical Company

10. What do you like best about EA?

Excellent technical services group. Very detailed explanation of data.

Director, Cell Genesys, Inc.

Comments:

Overall, our experience with EA has been nothing but positive

Senior Research Scientist, Incyte Corporation

CONFIDENCE

For more information: 866-293-6094

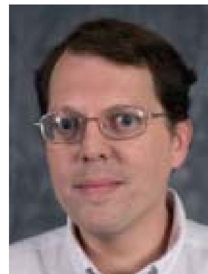
Expertise Beyond Expression℠

Markers NEWS

Supercomputing: New Processor Architecture Holds Promise for Protein, Gene Studies

A supercomputing architecture that first appeared in prototype form more than 10 years ago has been given a new lease on life, thanks in part to a recent $4 million Department of Defense grant issued to seed the new Center for Adaptive Supercomputing Software. The joint project teams up Pacific Northwest National Laboratory and supercomputer maker Cray, as well as several institutions including Georgia Institute of Technology and Sandia National Laboratories.

The initiative aims to develop software that takes advantage of the multithreaded processing capabilities of Cray's XMT supercomputer. Unlike traditional supercomputer processing architecture, where each processor gets a portion of memory for each calculation in a piece-by-piece fashion, the new processors are each capable of multiple, simultaneous data crunching and use a much larger pool of memory per processing core. This design means that many disparate sets of complex data can be digested at once, instead of each portion of data being handled piece by piece.



DAVID BADER

David Bader, a computer scientist at Georgia Tech, demonstrated the architecture's application to biology by identifying proteins that, when knocked out, disrupt the cancer-causing networks in a particular cell. Bader and his team used a social networking algorithm to mine a huge collection of publicly available human proteome datasets. "This is similar to finding important people in a social network, sometimes called 'connectors,'" Bader says. "Looking for these proteins is like looking for a needle in a haystack — and it is usually computationally intensive that won't work well on current [high-performance] machines, but this new architecture is really designed for this type of problem."

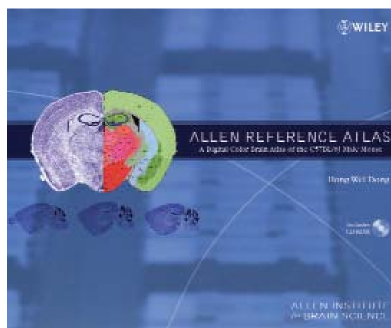
Normally, multiple database searches of this kind would take hours and hours to complete on a typical cluster or supercomputer. But with an algorithm specially ported to this multithreaded processing architecture, the same job takes mere seconds to complete, says Bader. "These sorts of problems have overwhelmed modest size clusters, and if you start adding processors to a cluster, it takes longer and longer to run because the communication costs dominate," he says. "This is really the first architecture where you can pose a biological hypothesis, test it out, and run it in short seconds or minutes versus hours to days, or maybe never."

Bader and his colleagues believe the concept could offer a lot to large-scale life science computing problems. "I think as we gather more genomics data we can use such a system to make scientific discoveries," he says. "I would hope, looking at three to five years, that we keep investing in these novel types of architecture and looking at the scientific results we can achieve, especially in the areas of solving genomic problems and understanding of the genome."

— Matthew Dublin

Book Review

The Allen Reference Atlas: A Digital Color Brain Atlas of the C57Black/6J Male Mouse, by Hongwei Dong and the Allen Institute for Brain Science



In this atlas, authors Hongwei Dong and the Allen Institute for Brain Science have created a full-color brain atlas of Nissl-stained fresh-frozen tissue from young adult male mice. To be used in complement with the Allen Brain Atlas gene expression database, the atlas includes coronal and sagittal sections and color-coded brain structure delineations. It comes with a free CD-ROM, which can be used in conjuction with online graphical, gene expression, and informatics tools.

Publisher: John Wiley & Sons, Inc.
Publication date: January 28, 2008
ISBN: 978-0-470-05408-6

— Jeanene Swanson

NEWS Markers

Clinical labs: AACC Annual Meeting Covers Genomics, Proteomics, and Diagnostics

At the American Association for Clinical Chemistry annual meeting held in Washington, DC, at the end of July, it was the changing role of clinical laboratory medicine in improving healthcare that garnered the most attention. In one of the first plenary sessions, Roy Vagelos, retired chairman and CEO of Merck, talked about the changing pharmaceutical industry and its evolving role in providing healthcare. One symposium followed up on that with a look at evidence-based medicine, while another delved into the challenges facing clinical laboratory testing in the developing world.

James Hughes, director of Emory University's program in global infectious diseases, took on some of the challenges and opportunities in creating diagnostics for the developing world. Those included managing the global burden of infectious disease, dealing with emerging microbial and vector-borne threats that have cropped up in the past decade, and implementing disease control efforts. Some of the most devastating diseases globally in terms of mortality are HIV/AIDS, tuberculosis, and malaria, Hughes said, citing data from a 2003 World Health Organization study. "What you don't hear much is that, actually, the leading infectious disease killer worldwide is lower respiratory tract infections, primarily pneumonia and influenza, and that coming in third is diarrheal disease," he added.

At the other end of the Walter E. Washington Convention Center, Donald Baldwin, director of the microarray facility at the University of Pennsylvania School of Medicine, spoke about the use of standard genomics and proteomics tools for molecular diagnostics and high-throughput genotyping, especially for cancer. While his talk focused on miRNAs, he touched on many up-and-coming research tools that could be utilized for functional genomics, including alternative RNA splice arrays, tissue microarrays, cell transfection microarrays, and "next-next-gen" sequencing, which uses nanotechnology to get down to the single-molecule level. "Functional genomics has long way to go," he said, citing headway made in proteomic technologies as just the start. "We need good ways of looking at protein-protein interactions and sub-cellular localization to really understand in this entire model, what's happening to the whole genome context."

A debate on whether warfarin testing was ready for primetime was also the subject of a full-day symposium, with a series of experts in the field giving both supporting and opposing arguments. While some say using genetic tests to determine warfarin dosing is ready to lead the way for more pharmacogenomic tests, others say the data is not yet available and that testing could increase costs while not offering any real benefits. Inside the exhibit hall, ParagonDx paired with DNA Genotek to offer a warfarin demo to conference attendees, including a test for CYP2C9 and VKORC1, the two variants that determine warfarin sensitivity. The simple procedure consisted of signing a few consent forms, spitting into a tube, and then going to a designated website where the results were posted.

Finally, to put it all in perspective, veteran broadcast journalist and political commentator for ABC News, Cokie Roberts, gave a lively plenary on the healthcare debate — what the US presidential election will bring in terms of the candidates' ideas for healthcare reform. "It is definitely going to be robustly debated in this election campaign," she said. "It's a different debate from what it used to be." She contended that this election will see people demanding healthcare reform, and the statistics are enough to explain why: from 2001 to 2007, worker earnings in the US went up 18 percent, while the cost of health insurance premiums increased by 78 percent, Roberts said. More than 25 percent of people say they have trouble paying for healthcare and insurance and "with serious illness, it makes it considerably worse," said Roberts. "For those who are fighting cancer, which now is one in three of us, 25 percent … said the disease had used up much or all of their savings. I think that the pressure on the national government to act is going to be very strong."

"We need good ways of looking at protein-protein interactions and sub-cellular localization to really understand ... what's happening to the whole genome context."

— Jeanene Swanson

Zeitgeist | BLOGOSPHERE BRIEFS

Something to Talk About

The blogosphere has been chattering about job prospects, George Church, women in science, and Charles Darwin.

By Ciara Curtin

Saving Your Job

With the uncertain economy, there have been rumblings about layoffs and takeovers. At In the Pipeline, Derek Lowe has a bit of job news: he heard that there will be chemistry layoffs at Pfizer come the fall, and that a potential Roche takeover of Genentech has gotten the latter's employees to look elsewhere for gainful employment. Lowe also doles out some advice on how to keep your job from being sent abroad: start generating ideas and doing more difficult chemistry. "You have to bring something that can't be purchased so easily overseas," he writes.

http://pipeline.corante.com

Be Whatever You Like

New York Times blogger John Tierney examines applying Title IX, the law banning sexual discrimination in education, to the sciences. The National Science Foundation, NASA, and the Department of Energy have been looking for instances of sexual discrimination at universities receiving federal funds by examining lab space and interviewing researchers. Critics worry this will lead to a quota system for women in science. A blogger at Adaptive Complexity writes that having quotas "would certainly make women second-class citizens in science, because they could never be judged on their own merit."

http://tierneylab.blogs.nytimes.com/

www.scientificblogging.com/adaptive_complexity/blog

The Summer of George

George Church exploded onto the mainstream scene recently. A profile in *Wired* shows the winding path that his career has taken. (That same article, as noted by the *Genetic Genealogist*, unmasked the Personal Genome Project's tenth participant: Harvard psychologist Steven Pinker.) In an interview with Charlie Rose, Church says that people should initially have low expectations of personal genomics. Valleywag also notes that one of Church's many advisory roles is with 23andMe, and that one of the co-founders of that company, Anne Wojcicki, is married to Sergey Brin of Google, which has backed Church's PGP.

http://www.thegeneticgenealogist.com/

http://valleywag.com/

Deleting Darwinism

With anticipation mounting for the 150th anniversary of Charles Darwin's theory of natural selection (and his 200th birthday), the man himself takes center stage. A blogger at 3 Quarks Daily links to videos of Richard Dawkins discussing his hero. But it's a suggestion of Olivia Judson's that catches on; she wants to get rid of the term "Darwinism." The 3 Quarks Daily blog jumps on the bandwagon, as do Evolgen, Larry Moran at Sandwalk, and Mike the Mad Biologist. "Modern evolutionary biology has gone way beyond Darwin's original ideas and it's no longer appropriate to describe the modern ideas as 'Darwinian,'" writes Moran.

http://judson.blogs.nytimes.com/

http://3quarksdaily.blogs.com/

http://scienceblogs.com/evolgen

http://sandwalk.blogspot.com/

http://scienceblogs.com/mikethemadbiologist/

Our Genotyping Solutions are More Comprehensive...

Because Everyone Responds Differently.

Accurately Detect Biologically Important Polymorphisms with Roles in Drug Metabolism and Efficacy.
Genes that code for drug metabolism enzymes (DMEs) and transporters are important targets in pharmacogenetic and clinical research. Genetic variations within these genes can influence the rate of drug metabolism and impact drug effectiveness. Our TaqMan® Drug Metabolism Genotyping Assays collection targets over 2,600 SNPs, InDels, and MNPs residing in coding and regulatory regions in over 220 DME genes. Robust, highly reproducible, and easy-to-use TaqMan® Assay allelic discrimination technology provides unmatched specificity and enables accurate detection of variants in highly homologous DME gene families.

All TaqMan® SNP Genotyping Assay products—including DME, Pre-Designed, and Custom Assays—offer a simple, scalable workflow for fast data generation. Learn more at www.allsnps.com

For Research Use Only. Not for use in diagnostic procedures. Practice of the patented 5′ Nuclease Process requires a license from Applied Biosystems. The purchase of TaqMan® SNP Genotyping Assays includes an immunity from suit under patents specified in the product insert to use only the amount purchased for the purchaser's own internal research when used with the separate purchase of an Authorized 5′ Nuclease Core Kit. No other patent rights are conveyed expressly, by implication, or by estoppel. © 2008 Applied Biosystems. All rights reserved. Applera, Applied Biosystems, and AB (Design) are registered trademarks of Applera Corporation or its subsidiaries in the US and/or certain other countries. TaqMan is a registered trademark of Roche Molecular Systems, Inc.

Careers | PROFESSIONAL LIFE

Savvy Strategies in a Budget Crunch

In a tight funding climate, what's the best way to get a grant? Scientists are trying to increase their odds by submitting more proposals, adding senior PIs to their applications — and other choices that are just plain wrong. *By Meredith Salisbury*

Some people call it the "undoubling" — referring to the current federal funding situation, which after several years of budget increases has now seen a dramatic fall-off thanks to inflation and years of flat appropriations.

The funding crunch has led to a host of other problems: skyrocketing numbers of grant applications, plummeting success rates, increasing age of winning a first grant, and more.

Just as there are good methods for writing a grant application in general, there are strategies to follow when submitting a proposal in this kind of funding climate.

The natural tendency of anyone looking at dwindling odds is simple: place more bets. Review panels are inundated with proposals as scientists have taken to submitting significantly more applications than they normally would. While the strategy seems like a no-brainer, chances are, it's working against you, says Joanne Tornow, acting director of the division of molecular and cellular biosciences at the National Science Foundation. Funding agencies don't want to cut the dollar amount of the awards they're giving, so they're more likely to cut the number of awards granted. At NSF, Tornow says, if you submit several proposals that appear related, they'll be evaluated together — and

NORKA RUIZ BRAVO

The Scoop

Fewer proposals

Of course, while there's a risk that someone else will keep submitting extra grant proposals, getting anyone to stop the application spree has slim chances of success. But NSF's Joanne Tornow says that driving up the number of proposals can weaken your overall chances and that your worst, rather than your best, submission may have a greater impact on reviewers.

Targeted applications

One way to see how well your research idea would fit in with a program's guidelines is to see what kind of projects have been funded through that program in the past. Make use of agency grant listings, such as NIH's CRISP, to see what's been successful before. Apply to the programs that seem like a natural fit for your research concept, and skip the ones that lean in a different direction.

Make contact

There's no substitute for getting in touch with the program directors themselves. You'll get the most up-to-date information about the program and specific advice on the funding opportunity. It's also just a good idea to get to know the people at the funding agencies.

Don't play it safe

In a funding crunch, scientists tend to cut off the more radical ideas in favor of the safer, more run-of-the-mill proposals. Tornow says that's a mistake; her agency looks for research projects that push the envelope.

Collaborate for expertise

If your proposal includes a type of research that you haven't attempted before, it might help to find a collaborator who has expertise in that particular area. Just make it clear in the application who will be responsible for what.

Don't lean on established scientists

Younger investigators fearing they won't get grant funding may be tempted to add a more experienced scientist as a co-PI on the application. Norka Ruiz Bravo and Tornow agree that this can undermine your ultimate goal of establishing yourself as an independent investigator. Agencies tend to have special opportunities for scientists early in their careers, so first try your hand with those.

the weakest of the batch may reduce the overall enthusiasm of reviewers for any of your proposals.

A stronger approach, Tornow says, "is to be very targeted on the proposal that you write." To that end, do your homework: certain agencies fund certain kinds of research, and the programs within those agencies are more specific still. With limited funding, each program director has to make sure every choice aligns with the program's priorities. From NSF's perspective, "there is a need to build portfolios and to think about where our investment can have the greatest impact," Tornow says.

When you scope out a program announcement, do some digging to see what kinds of projects have been funded by that program (or a similar one) in the past. Does yours fit the scope and direction of those? If your proposal would be an outlier from the types of projects that have been successful, focus instead on finding one that's a better fit. Avoid the trap of playing it safe, though: Tornow says her agency looks for ideas "that are a little bit more out there" — which means "you need to guard against those tendencies to move toward the safe stuff."

Once you've found a program you're interested in, the best way to get the inside scoop is to contact the program director and ask questions. "The best thing that anybody can do is get on the phone and call us," says Norka Ruiz Bravo, director of the Office of Extramural Research at NIH. If cold calling intimidates you, try e-mail — it's unobtrusive and can be a great way to ask a simple question or two.

Young investigators nervous about funding tend to consider teaming up with a more established scientist to act as co-PI on the grant, figuring it will improve their odds. While collaborations can be a great idea, leaning on more experienced scientists can backfire. "I would advise someone to be independent as quickly as possible," says Ruiz Bravo. Even if the grant idea is your own, you run the risk of being perceived as riding on the other scientist's coattails.

Still, teaming up is a good idea in cases where you're proposing to do work you haven't been trained to do, says Tornow. In such a situation, find someone who has expertise in that area and add that person to the proposal. Even then, Tornow cautions, "it is going to be important for it to be clear whose intellectual input is driving the project."

Next-Gen DNA Shearing

Tune in High Definition Sample Preparation with Covaris™ Systems

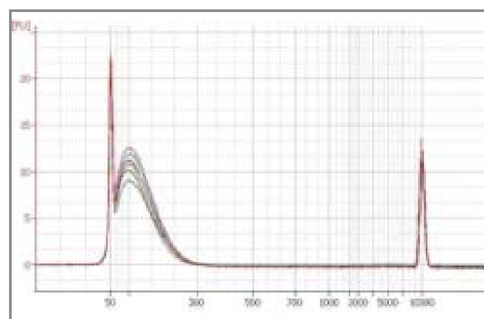
The **S2** is powered by the Covaris Adaptive Focused Acoustic (AFA) technology; a proprietary computer controlled, focused acoustic process based on shock wave physics.

Benefits include:
- Non Contact
- Rapid
- Closed vessel
- Isothermal
- Automatable
- Reproducible

Covaris™ Bringing High Definition Sample Preparation to Science™

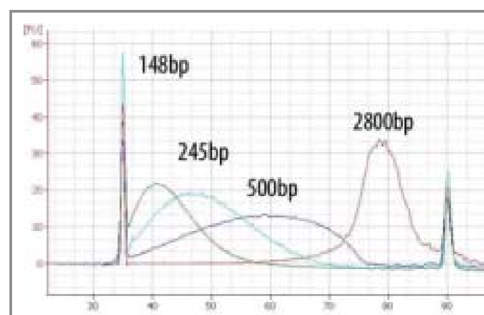
www.covarisinc.com

Reproducible



N=12
Avg=100bp
CV =2.6%
S2 and E210 systems
TE buffer
100ul volume
Lambda DNA
(starting material)

"Tune In" Size Range

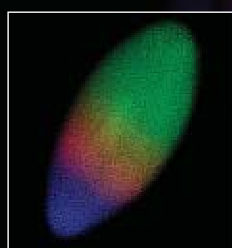


Dose
<5min
DC=20%
Int=5
CPB=<1000
Tris-EDTA
100ul
5°C

©2008 COVARIS, INC—ALL RIGHTS RESERVED.

Parlez-vous MATLAB?

Over one million people around the world speak MATLAB. Engineers and scientists in every field from aerospace and semiconductors to biotech, financial services, and earth and ocean sciences use it to express their ideas.
Do you speak MATLAB?



Quantitative high-throughput gene expression imaging using data from FlyEx Database.

This example available at mathworks.com/ltc

Image from FlyEx Database. Used by permission. http://flyex.ams.sunysb.edu/flyex and http://urchin.spbcas.ru/flyex

©2007 The MathWorks, Inc.

MATLAB®
The language of technical computing.

FRAN LEWITTER & GEORGE BELL Informatics Insider

Homology: Genealogy for Genes

Homology may be a keystone of biology, but confidently identifying homologs is a challenge. Here, we go beyond basic sequence comparisons to look at better methods to assign homology.

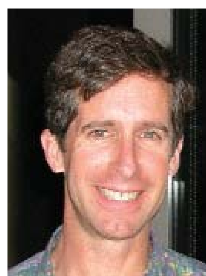
The great power of model systems in molecular biology has been apparent ever since early researchers used bacteria, yeast, worms, and flies to learn about the human body. More recently, the power of comparative genomics has been harnessing evolution to help identify the most obviously important parts of our DNA by linking a piece of one genome to a corresponding piece of another genome.

On a molecular level, both of these approaches can require that we link genes that are homologous, i.e., share a common evolutionary origin. Even in a junior high school biology class, one can very easily define homology as features (such as genes, proteins, or even structures) that arise from the same ancestral entity. Devising and applying an operational definition of homology that is practical and comprehensive, however, keeps a lot of biologists and bioinformaticians quite busy. Here we discuss some methods of assigning homology, along with some of our challenges that show why this is a problem that can't be effectively addressed with basic sequence comparisons.

Homology has been described by David Wake as "the central concept for all of biology." As bioinformatics people, we're often asked to identify the homolog of a human disease-causing gene in another species, whether that be mouse, zebrafish, yeast, or another model system. Just about any molecular biologist can now use Blast to take a human protein to search a database of, for example, zebrafish proteins to identify the most similar one. Is the top hit the homolog we're looking for? We can't be sure, and this gets at the crux of the definition; homology — or more specifically, orthology (separated by speciation) and paralogy (separated by gene duplication) — is a hypothesis that reflects a history of shared origin that can be supported but not unequivocally proven. We can quantify similarity between proteins or gene sequences using percent identity, length of alignment, or even domain structure, but we can't quantify homology; either features are homologous or they aren't.

Homology is commonly interpreted to mean present in the last common ancestor, so even if all proteins evolved from the same good bits of primordial soup, knowing this distant shared ancestry isn't so useful. To assign pairs of homologs A and B across species, genome-scale analyses often go at least one step further than our Blast search above, requiring that B is the most similar protein to A and vice versa. If we find an orthology pair like this, we're in good shape, but do we want to further restrict our measure of similarity? What if we can generate only a short local alignment? What if a similar analysis of gene sequences is inconsistent? What if different scoring matrices produce different results? We'll probably try to optimize the details of our homology search for the specific use of these data, but our homology assignments will still be open to debate. Also, unless this homology is well established, we'll want to make sure to explain our operational definition.

FRAN LEWITTER


GEORGE BELL

Digging deeper

What if our operational definition of homology doesn't turn up any orthologs of our favorite human gene using reciprocal Blast search? Has the missing gene just not been sequenced or annotated yet? Or is it actually missing from the genome

of our favorite model organism? How about if the human gene is present by name in the other species? All of these possibilities may need investigating. It would be much easier for us if orthologs had the same names in different species, but even some genes that do have the same names don't appear to be orthologous, now that we know more complete gene catalogs. Some of these cases fit into the unfortunately-named category of "functional homologs" which are proteins with similar functions but not of shared evolutionary origin (and therefore not actual homologs).

The most powerful current methods use information from multiple species at once, and this orthology determination benefits from the ever-growing number of genome assemblies.

It would be very convenient for biologists and database administrators if all orthologs were clearly 1:1 where, for example, one human gene is orthologous to one mouse gene. If analysis of the mouse genome shows that virtually all human protein-coding genes have mouse orthologs, then why is it so hard to link every human gene to a mouse gene? Gene duplication and subsequent divergence, giving rise to paralogs, can make determination of orthology much trickier. If we discover two obvious human paralogs and two mouse paralogs, all of which appear to be homologs, how can we figure out which mouse gene is the ortholog of each human gene? If gene duplications occurred after speciation, then there may not be any 1:1 orthologs.

These 1:many or many:many homology relationships create extra challenges for comparing genome-scale datasets across species. On the other hand, if it appears that a gene duplication event occurred before speciation, we can try to resolve multiple homologs into 1:1 orthologs. All of this can be done better now than ever before, in part thanks to improved genome assemblies and gene sets. This can reduce strange observations, such as a recent look at a collaborator's favorite gene in a fish. The fish genome assembly and gene prediction pointed to this gene's presence in a set of a couple dozen highly similar paralogs, a degree of gene expansion that was absent in other species. Further investigation led to the much less interesting explanation that the expansion of this repeat-flanked gene was very recent, having just occurred in the most recent genome assembly.

Biomedical researchers who use mammalian model systems have a much easier time identifying homologs of human genes than others who experiment on worms, flies, and yeast. First, the genes themselves have had much less time to diverge, so the orthologs are much more similar. Second, the genomes have had much less time to diverge, so chromosomes have much longer conserved syntenic blocks. As a result, if mystery gene B is flanked by genes A and C in human, each of which have clear orthologs A' and C' in mouse which are close to each other on the same chromosome, we can look between these mouse genes to try to find B'. This conserved surrounding genome environment is stronger evidence, in addition to protein and/or gene similarity, that genes are really homologs and not just similar genes. On the other hand, alignment of genomes is not a solved problem, and alignment gaps do not always mean lack of homology.

The most powerful current methods use information from multiple species at once, and this orthology determination benefits from the ever-growing number of genome assemblies. The Ensembl project, for example, leverages the power of comparative genomics by using Blast to search with each gene against all other genes (species by species), clustering the similar sequences, building multiple sequence alignments, and then generating phylogenetic trees which can be compared to a species tree. The use of sequences at different phylogenetic distances helps resolve a lot of cases that would be difficult to figure out with only a pair of species at a fixed distance. This brings up the final way to determine homology: consult a reliable database that has already done the best possible large-scale homology analysis. These databases aren't foolproof, but they're a great place to start.

Just as with other scientific statements, we shouldn't believe everything we read or hear about an inferred homology relationship. If they're important to us, we probably need to investigate the genes further so we can hopefully convince ourselves and others: either they are homologous or they aren't.

Fran Lewitter, PhD, is director of bioinformatics and research computing at Whitehead Institute for Biomedical Research. This column was written in collaboration with George Bell, PhD, a senior bioinformatics scientist in Fran's group.

Sweet Time for Informatics

Following in the footsteps of genomics and proteomics, the budding field of glycomics is starting to show signs of getting its informatics act together. *By Matthew Dublin*

While the nascent field of glycomics has not received nearly as much attention and funding as its more established systems biology siblings (namely, genomics and proteomics), this small but steadily growing research community is solving its own database and software challenges.

Just like genomics and proteomics, glycomics has an "ome" as its holy grail — although this one is a long, long way off from completion. The glycome is a complete map of all the complex carbohydrate or glycan structures in a particular organism. These intricate sugar structures have been shown to play a key role in everything from pathogen recognition and sperm-egg interaction to immune system response. In addition, many glycoproteins have been identified as biomarkers for cancer and several diseases.

Unlike DNA, RNA, or proteins, which are all template-driven, glycans are created by the actions of a large number of enzymes, which can result in a seemingly endless number of structural variations. "If someone wants to decode the glycome of a cell, it's a fairly complicated process. It's not the same as someone saying, 'I want to decode the genome or proteome,'" says Rahul Raman, director of the bioinformatics core for the Consortium for Functional Glycomics. "In a cell, you have different glycoproteins, and each protein has multiple glycosylation sites. At each site, you have a variability of the type of glycans that can be expressed, so you can see how complex it becomes when you want to know every glycan at every glycosylation site of every glycoprotein in a particular cell."

Glycan databases

Compared to both the availability and sophistication of databases and software tools for genomics and proteomics, glycomics trails way behind. And just as each of those has gone through its own database growing pains, glycomics must first get its data resources up to speed before more researchers and commercial vendors have a reason to start seriously contributing software tools to this area. Currently, there are three major databases that house glycan structure data: KEGG Glycan, Glycosciences.de, and a relational database hosted by the Consortium for Functional Glycomics, an international initiative funded by the National Institute of General Medical Sciences. The CFG's database is a Web portal connected to the integrated interfaces of diverse datasets in the CFG's relational databases, which contain content on glycan-binding proteins, glycan structures, and glycosyltransferases. KEGG Glycan is an extension of the Kyoto Encyclopedia of Genes and Genomes database and is managed and developed by the Kyoto University Bioinformatics Center. Glycosciences.de is maintained by the German Cancer Research Center and provides researchers with mass spec and glycan structure data as well as applications for glycan analysis.

While certainly not impressive to those familiar with current proteomics and genomics databases, these repositories do mark an important development for glycomics. Prior to their arrival, the only major glycan structure resource available to researchers was CarbBank, a database hosted by the University of Georgia that served as the de facto central repos-

> **"If someone wants to decode the glycome of a cell, it's a fairly complicated process. It's not the same as someone saying, 'I want to decode the genome or the proteome.'"**

itory for all glycan structures. CarbBank had its heyday in the 1990s and has since run out of funding, although it provided a large part of the glycan structure data and system architecture for two of the three newer databases.

Though the three current databases share the same initial collection of glycan structure, they use different file formats, a huge informatics stumbling block. KEGG Glycan uses its own KEGG Chemical Function Format; Glycosciences.de uses the LINUCS format; and CFG uses a format established by the International Union of Pure and Applied Chemistry.

"A lot of [glycan] structures are still not represented among all these databases, and it will take time and money for a repository like GenBank for glycobiology."

Setting standards

In September 2006, a workshop was held at the National Institute of Health so that glycobiologists from across the globe could assess bioinformatics needs and the current state of glycan structure analysis tools. An outgrowth of this meeting was the establishment of a standard file format for exchanging glycan structure data. They chose the GLYDE-II XML file format, developed by William York, an assistant professor at the Complex Carbohydrate Research Center at the University of Georgia. And while this is certainly exciting to many, it's like a bunch of TV owners still using bunny ears learning about HDTV. "I think it's fantastic that the format was agreed on, and that's really going to help," says David Goldberg, a research fellow at the Palo Alto Research Center. "But it's not used much — not because there's anything wrong with the standard itself, it's just that there is not that much software out there yet that's designed to take advantage of it."

Encouraging glycomics researchers to adopt a standard file format for glycan structure data submission would be beneficial not only to facilitate independent database integration, but also to make incorporating experimental data published in journals easier. "Each individual database has made [its] own attempts to update their data according to the literature, but it's hard because of the variety of notations used to represent glycan structures," says Kiyoko Aoki-Kinoshita, an associate professor of bioinformatics at Soka University in Tokyo. "In general, it can be assumed that a lot of [glycan] structures are still not represented among all these databases, and it will take time and money for a repository like GenBank for glycobiology to be developed."

Aoki-Kinoshita and others believe that the most urgently needed improvement is the consolidation of these databases along with supplementary data such as pathways, interacting proteins, and binding affinity into a one-stop resource. The creation of such a resource was also deemed a priority at the 2006 NIH meeting, as leaders in the field hope that a standardized glycan structure data file format such as GLYDE-II XML will eventually lead to a centralized and curated glycan structure database.

"There's really a need to get the scientific community and the journals to agree on certain guidelines, so whenever someone wants to deposit a structure they could just do it through a central submission system — and then that structure will automatically go to the different large initiative databases," says Raman. "All of us right now are trying to manually collect this information because there is no system to deposit a structure that will automatically be piped into the different databases. That's the main challenge in maintaining and expanding the current glyco-databases."

Early tools

Still, serious gains have been made since the formation of the CFG and other large-scale initiatives geared toward mobilizing the glycomics community. But Goldberg says that it's a bit difficult to predict how long it will take for glycomics to catch up to proteomics and genomics in terms of software development. Many in the field feel that this is due partially to the fact that glycomics is still too small a sector of the market for commercial developers to care about, although a handful of vendors have started offering some tools. Proteome Systems has a glyco-database and a suite of software tools for analyzing mass spec data and structure prediction. And Premier Biosoft International is also pitching SimGlycan, its mass spec software analysis tool geared toward studying glycosylation, a key area of post-translational study for glycobiology that looks at when glycans attach themselves to proteins. Still, many glycobiologists agree that, for now, most of the software development will come from academia.

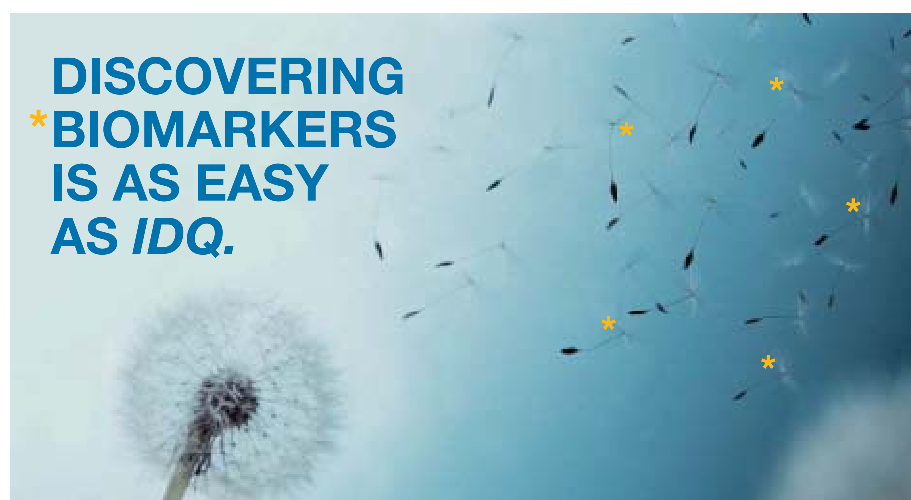
Along these lines, Goldberg has developed an automatic annotation software tool called Cartoonist that works with single MS data to determine the composition of a particular glycan structure. The program works by selecting the most plausible annotations for each peak in a mass spectra profile from a library of possible cartoons. Goldberg says the current version of Cartoonist is unique among software tools; earlier research in glycomics utilized MS/MS because researchers were merely copying the same techniques that worked in proteomics.

"From single MS data, Cartoonist lets you figure out what the glycans' compositions are, and then it makes a very good first guess at what the actual structures [are]," says Goldberg. At the moment, those wishing to use Cartoonist must send their spectra directly to Goldberg, but he says that will change in the next year as he works out the kinks and beefs it up to include MS/MS data. He hopes to distribute the tool to CFG members, and then ultimately to make it more widely distributed. "The tools used to assist in the annotation of glycan mass spectra have made major contributions to this field," says Aoki-Kinoshita. "In particular, the Cartoonist suite of software, which is being used by the CFG to annotate the large amount of data they are generating, has been apparently very useful."

Over at the CCRC, GlycoVault, a Web-based informatics gateway that contains databases, ontologies, and other glycan structure-related data, is also promising. GlycoVault is hosted by the University of Georgia's Integrated Technology Resource for Biomedical Glycomics, an initiative funded by the National Center for Research Resources. This application also contains the Glycomics Browser, a Web-based visualization and analysis tool for glycan data.

Overall, the lack of software and the glyco-informatics community's small size may be a sort of vicious cycle hampering its growth. "It's a chicken or egg problem. Because the [glycomics] community is small … there isn't much of a demand for software," Goldberg says. "But on the other hand, if there was better software, maybe more people would do this kind of experiment, so I think it's going to ratchet up slowly." This leaves researchers like Goldberg and others with the onus of providing tools to the small but growing community.

And as was the case in the early days of genomics, once that divide between computer scientists and bench biologists is traversed and the databases become more developed, things will ramp up. "I think that once the data can be accumulated, the bioinformatics fields can start to develop methods specifically for glycobiology, but it is important for informaticians to work closely with experimentalists in order to develop useful tools," says Aoki-Kinoshita. "The language barrier between bionformatics and glycobiology needs to be broken down [and] it is my hope that this can be overcome in the near future."



TargetIDQ™ Services

Defining the future of Metabolomics

Biocrates' services enable you to identify and quantify metabolomics biomarkers in a broad range of fluids and tissues. Our assays are robust, reproducible, and use the least amount of sample in the industry.

Metabolomics Society, Boston, MA
September 3-6, 2008

Optimizing Cell Culture Development, Boston, MA
September 15-16, 2008

CREATING KNOWLEDGE
FOR HEALTH

sales@biocrates.com
www.biocrates.com

Starting from Scratch

Ken Stuart founded the Seattle Biomedical Research Institute with a single grant and high hopes. Thirty years later, a 260-member staff and a $40 million budget suggest he did something right. *By Meredith Salisbury*

Let's say you won a grant worth $140,000. What would you do with it? Chances are, parlaying that funding into your own fledgling research institute would be pretty far from your mind.

But that's just what Ken Stuart did in 1976, when he was in his mid-30s and received his first NIH grant. It was a modest, two-year award at $37,000 per year (adjusted for inflation, that's worth about $140,000 today). He used the money to rent space in an office park, start up his research lab focusing on global infectious disease, and pay his salary. It was a humble beginning for the Seattle Biomedical Research Institute.

SEATTLE BIOMEDICAL RESEARCH INSTITUTE

Stuart says that his early training experiences had provided examples, both good and bad, of different types of research environments. A stint at the National Institute for Medical Research just outside London showed him an institution that "enhanced the ability of scientists to conduct research," while a teaching gig at a Florida university made clear the distracting duties of teaching and other administrative necessities. His vision was simple: to create a "research environment that was efficient and effective" — a place where research was first and foremost, and any other tasks were kept to a minimum.

Stuart's gamble paid off: "People were attracted by the environment that I'd created," he says, and over the next decade the institute came to be known as a place to do intensive science. But with that emphasis on science came a lack of overall goals for the institute as a whole. "The way the institute developed, it really was sort of a collection of independent laboratories," he says. By the '90s, Stuart had become interested in pursuing infectious disease research beyond the basic lab and wanted to focus on translational work. Internal discussions over time revolved around whether to remain as a series of "boutique laboratories," says Stuart, or to make the transition to "a full-fledged research institute with an integrated vision and mission." They made the transition.

In early 2004, SBRI moved into a new $40 million facility, built after a major fundraising effort. Today, the institute occupies some of the space and leases out the rest to researchers from a local children's hospital as well as some companies. Stuart says that he expects the institute, which has been growing at about 25 percent annually for the past five years, to eventually take over the entire building. Today, SBRI boasts a staff of 260 with 15 principal investigators; Stuart plans

to increase that over time to 400 full-time employees.

Need for translation

While getting scientists to buy into the vision of one institute instead of many labs might have taken some discussions, Stuart says the real challenge in the course of the group's development has been getting people to embrace translational research. At the very beginning of the institute, this wasn't even an option, he says. In the mid-'70s, studies on infectious disease were so early-stage that "I could not really see a way to do research that would support the activities that would lead to interventions," Stuart recalls.

But in the past decade or so, that has changed radically, especially in the areas that SBRI scientists focus on: HIV, tuberculosis, malaria, and emerging infectious disease. The real problem was that the public funding system "didn't reward" translational work, Stuart says; scientists are rewarded for basic research and for publishing. "We had many discussions about [working toward an intervention] — why that's strategically important, why it's morally important."

Today, SBRI scientists are months away from clinical trials for a malaria vaccine, and they have other promising candidates for HIV and malaria in the pipeline. The first one, which is expected to enter trials in 12 to 18 months, began with work by malaria expert Stefan Kappe, who joined the institute five years ago. He says a significant lure of SBRI was that as an independent institution — rather than, say, a program within a medical school — it's "very nimble" and "we can shift our priorities very rapidly."

Kappe knew that working toward a vaccine was part of the goal of the malaria program, so he focused his research on the liver stage of the parasite, which is when it enters its human host but remains undetected for seven days before spreading out into the bloodstream and causing infection. Using functional genomic studies and mouse models, Kappe and his team were able to determine the gene expression pattern during this stage for the first time. Knockout studies proved that deleting certain genes at this stage left the parasite unable to complete its life cycle and infect the host. In mouse models, dosing the mice with this genetically engineered version of malaria induced a "very powerful protective immune response" — such that when the mice were then infected with full-strength malaria, the mice proved to be completely immune over the course of their lifetime. During the upcoming clinical trials, the same process will be performed on human patients to see if the immune response holds true.

Another potential malaria vaccine comes from Patrick Duffy's group, which is studying pregnancy malaria. The team has identified a key protein used by parasites in the placenta, and are aiming to develop a vaccine to target the protein.

>SEATTLE BIOMEDICAL RESEARCH INSTITUTE
Seattle, Wash.

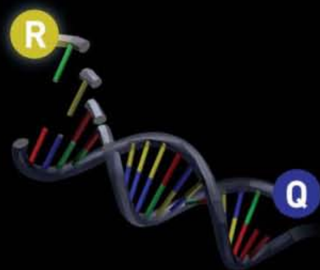
DIRECTOR: Ken Stuart
ESTABLISHED: 1976
SIZE: In March 2004, the SBRI team moved into a new $40 million, five-story facility. The institute currently leases out space to researchers from Children's Hospital & Regional Medical Center and to companies, but planned expansions to the staff will likely have the institute take over the whole building in the future.
STAFF: During the past five years, SBRI has been growing at about 25 percent annually, bringing its headcount to 260 employees (not all of whom are full-time). The institute plans to increase that to 400 full-time employees. There are 15 PIs on staff.
FUNDING: The institute's current annual budget is $40 million, almost all of which comes from grant funding, though the SBRI team has begun to focus more on fundraising with philanthropic organizations.
FOCUS: The main program areas are HIV, malaria, tuberculosis, and emerging infectious disease.
CORE LABS: SBRI's cores include technologies for DNA sequencing, imaging, protein production, proteomics, bioinformatics, and more.
COLLABORATORS: SBRI places a strong emphasis on partnerships with other institutions to help extend its reach and capabilities. The group has more than 100 alliances, with partners including Microsoft, Harvard, Walter Reed Army Institute of Research, the Pasteur Institute, and Novartis, to name a few.

Going global

As he was getting the institute going, Stuart realized that his group would never be large enough to have expertise in all areas of infectious disease — so collaborations have been essential to SBRI's development. Currently, for instance, many of the partnerships help the institute connect with people who are experts in developing vaccines. The institute's website lists 125 collaborations with partners ranging from New York University to Novartis to the University of Nairobi. "From our total budget," Stuart says — this year, that's $40 million — "about half of it goes out the door to support our collaborative activities."

PROBES DON'T HAVE TO BE PAINFUL...

...especially when you order custom TaqMan® dual-labeled probes from Operon Biotechnologies.



$95

Custom TaqMan® probes

- FAM-BHQ1 or FAM-TAM
- 50 nMol synthesis
- HPLC purification
- 100% QC by mass spec

Go to **www.operon.com** and use promo code GotProbes.

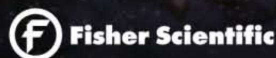
Free

Alien T-shirt

Go to **gotprobes.com/gt** to get your FREE Alien T-shirt.

operon
oligos · sequencing

www.operon.com
1-800-688-2248
oligo-us@operon.com



For customer service, call 1-800-766-7000.
To fax an order, use 1-800-926-1166.
To order online: www.fishersci.com

Shipping not included in promotional price from Operon. Offer valid only in USA (except in Roswell, New Mexico) and in Canada through December 31, 2008. TaqMan is a trademark of Roche Molecular Systems, Inc.

METABOLOMICS

Taming Metabolites

Metabolomics studies inch scientists ever closer to understanding phenotype. But to really make progress, pioneers are working on improving the technology and analytical tools of the field.

BY CIARA CURTIN

A 1966 *Biochemical Journal* article recently unearthed by Gary Siuzdak's lab at the Scripps Institute looks to be the first metabolomics experiment — though they certainly didn't call it that then. In it, researchers from Baylor College of Medicine describe using gas-liquid chromatography to separate metabolites from urine and tissue extracts. They also add that GC-coupled mass spectrometry gives a "diagnostic tool of great power" — something today's researcher already knows. But also in that article, C.E. Dalgliesh *et al.* grumble about overlapping peaks, resolving those peaks, and the lack of a database housing known metabolites. Sound familiar?

Today's metabolomics researchers have the same gripes, but are better poised to do something about them. The technology isn't very fast or robust as compared to other big players in systems biology, and identifying all the metabolites in a sample can be nearly impossible. At each step, from sample preparation to data analysis to metabolite identification, metabolomics as a field is still trying to resolve robustness, identification, and analysis issues that the other disciplines within systems biology have already overcome. Metabolomics, though, is working hard to live up to its potential and join the ranks of the high-throughput fields. Researchers are capitalizing on the track record of NMR and the sensitivity of mass spectrometry to increase the number of metabolites detected and then identified through new databases.

GARY SIUZDAK

"When it comes to metabolomics, it's just a plethora of different techniques and technologies and protocols, and some it is reproducible and some of it isn't. It's a bit of a Wild West, but there's a bit of consolidation and the trends are there," says the University of Alberta's David Wishart, who heads up the Human Metabolome Project. "I think in a year or two it will be a more mature field with more consolidation and more consistency and more robustness."

Detection

In metabolomics, the divide has been between nuclear magnetic resonance and mass spectrometry. Both tools can identify a sample's metabolite population to varying degrees of success. More and more, though, researchers are combining the approaches to take advantage of their strengths while minimizing their weaknesses. At the same time, people are using new tools and methods for both separation and detection to take a gander at their metabolome of choice.

NMR, which dates back to the 1940s, has long been used by chemists to identify molecules in a sample. As a tool, it gets kudos for being stable, reliable, and robust. "In NMR, you can analyze the same sample today and this time next year and get a very similar result," says Warwick Dunn at the University of Manchester.

The tool has a variety of roles in the lab. For one, it can be used to get a first look at what the metabolome contains. "NMR is our main technique, which we would use first as sort of a survey technique," says John Lindon, a professor at Imperial College London.

Or it can be used for metabolite profiling. "NMR is very good because it tells you exactly which position in a molecule contains a ^{13}C or ^{15}N, whereas mass spec only tells

Things to consider before purchasing a **Next Gen System.**

SCALABILITY
Throughput advances without costly upgrades

THROUGHPUT
More than 6 GB and 240M tags per run for cost-effective, large-scale resequencing and tag applications

FLEXIBILITY
Two independent flow cells to run multiple applications in a single or staggered run

ACCURACY
System accuracy of greater than 99.94% for the power to detect rare, causative SNPs without high false positive rates

MATE PAIRS
Libraries with inserts ranging from 600 to 10,000 bp for resolution of structural variation

DATA SETS
Publicly available human datasets for accelerating biological discovery and analysis tool development

SUPPORT
Trusted service and support team available when and where you need them

Applied Biosystems

you how many positions are labeled, not which ones," says Andrew Lane, a professor at the University of Louisville's James Graham Brown Cancer Center.

And, of course, NMR has its downside. "The big knock about it is that it's not very sensitive," says Wishart. Indeed, according to Wishart and Siuzdak, an NMR-based characterization of a tissue sample or biofluid yields a little more than 50 molecules, but looking at that same sample with mass spec methods can yield hundreds or even thousands of molecules, depending on the chromatography technique coupled to the mass spec. Some scientists use NMR for surveying, and then apply mass spec for a more targeted analysis.

Manchester's Dunn focuses on mass spec — particularly liquid-chromatography mass spec — and he works on developing and optimizing methods to use it in metabolomics. While mass spec may be able to see more metabolites than NMR, it has its own drawbacks, primarily reproducibility. In his lab, Dunn says, two separate sample sets might give 50 interesting metabolites, but only 10 of them overlap.

"In an ideal world, you'd use both technologies because, in any analytical technology, there is some bias in what it can detect, whether it be the type of metabolite it can detect or the sensitivity, for example," Dunn says.

In particular, he works on increasing the reproducibility of mass spec by using an automated closed-loop strategy that has minimal human intervention. Over many iterations, the Robot Chromatographer, as his team calls it, initializes the instrument settings and then changes them as it cycles through looking for the optimal settings. When used on GC-TOF mass spec, Dunn and his colleagues increased the number of peaks seen by three-fold.

Newer technologies are also coming onto the scene to topple NMR, LC/MS, and GC/MS from the top spots in metabolomic technologies. "The mass spec technology is wonderful now. The robustness is great, especially the new time-of-flight and quadrupole time-of-flight mass spectrometers. They have improved dramatically in the last couple of years," says Siuzdak.

Not only is the detection step being improved, but advances are also coming along on the separation side. Ultra high-performance liquid chromatography came on the scene a few years ago, using higher pressures and smaller particle sizes to increase resolution and sensitivity, allowing scientists to detect even more metabolites. "The more things you can see, the greater overview you can get of the biology of the system," Dunn says. He's not the only one looking into UPLC: Lindon and his group have begun to couple it with time-of-flight mass spectrometry.

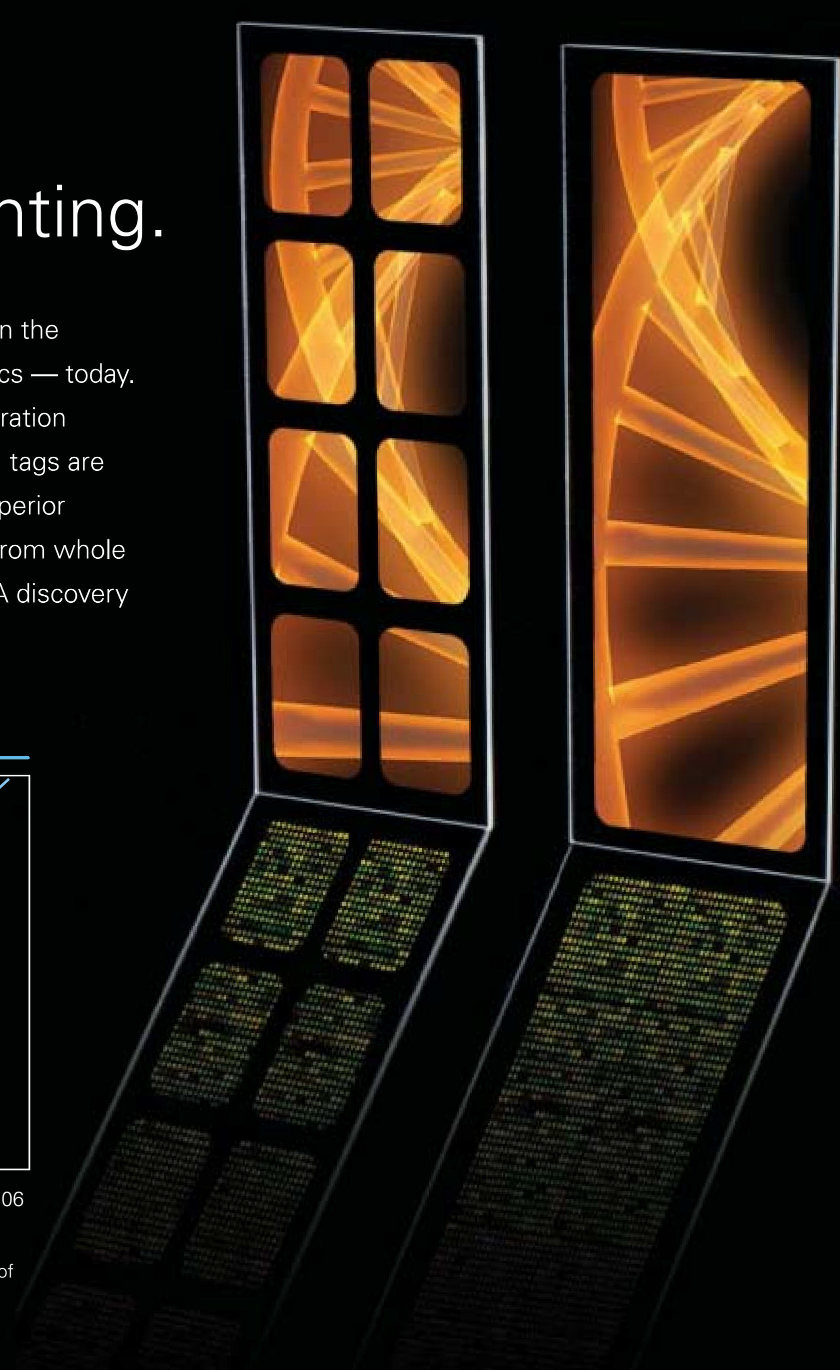
Another separation approach that is catching on is HILIC. Hydrophilic interaction chromatography allows researchers to detect more of the small, hydrophilic molecules that often are removed during a wash or that come out at the very beginning of the separation. Siuzdak is particularly intrigued by this method. "I've been recently surprised by some of the results that we've been getting that's allowed us to certainly see new things," he says. He is currently using HILIC to try to detect new molecules in knock-outs. "It's just another window into these samples," Siuzdak says.

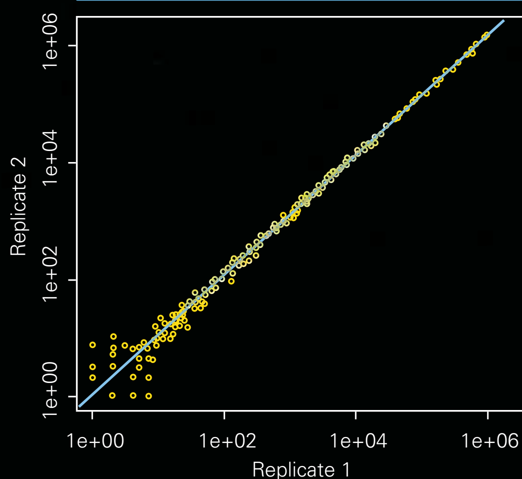
Deconvolution

The data that comes out of the end of NMR or mass spec is a mess of peaks and spectra. Making sense of all that can require some serious analysis, though some old hands can recognize NMR peaks just by looking at them. Most scientists rely on software packages to resolve the curves and deconvolute the data into something resembling a list of metabolites. "If you use chromatography on very complex

240,000,000 tags and counting.

Only the SOLiD™ System delivers on the promise of next generation genomics — today. With a unique two flow cell configuration and demonstrated scalability, 240M tags are just the beginning. The system's superior flexibility supports studies ranging from whole transcriptome analysis to small RNA discovery — the SOLID Generation delivers.





The SOLiD™ System demonstrates reproducibility of 99.6% and a dynamic range of six logs.

For SOLID PROOF, visit solid.appliedbiosystems.com

For Research Use Only. Not for use in diagnostic procedures. © 2008 Applied Biosystems. All rights reserved. All other trademarks are the property of their respective owners. Applera, Applied Biosystems, AB (Design) and ABI PRISM are registered trademarks. SOLiD is a trademark of Applera Corporation or its subsidiaries in the U.S and/or certain other countries.

samples like urine or plasma or something, there would be so many small molecules and metabolites eluting at pretty much similar chromatographic times that you get overlapped peaks, which makes them difficult to quantify and it makes it difficult to identify what it was," says Henrik Antti, an associate professor at the University of Umeå in Sweden. Different research groups are developing new and better software to help deconvolute what's in a metabolome.

At Imperial College, Lindon and his colleagues have developed and are using a statistical analysis method to identify NMR peaks. "It's not like a gene chip where you have one spot equals one gene," he says. "Here, a molecule, a metabolite will give many peaks on the NMR spectrum. We can use what we know about NMR to identify where those come from."

Building on a previous tool, called TOCSY (for total correlation spectroscopy), Lindon's team made a tool called STOCSY. This new method takes advantage of the correlation between peaks in NMR spectra — that multiple peaks can come from the same molecules and always occur in proportion. As an example, Lindon points to lactate, which has two NMR peaks — one from the methyl group and one from the CH group. Since NMR detects the hydrogen atoms of these groups, these two peaks will always be in a proportion of three to one. "We can use that statistical correlation to prove those two peaks are linked across hundreds or even thousands of samples," Lindon says. This relationship can help researchers work out which peaks of an NMR spectra go with which and, Lindon adds, help them identify potential biomarkers.

At Scripps, Siuzdak and his colleagues developed their own tool to analyze mass spectrometry data for metabolite profiling. Their XCMS is an open-source data analysis software package for LC/MS data that not only peak-picks, according to Siuzdak, but uses endogenous metabolites found in all the datasets as internal standards and aligns the peaks based on the retention time. Then, XCMS looks through its analysis and find the peaks that change between the dataset that are statistically relevant. "So now you have a set of molecules, typically, that look very interesting," says Siuzdak. He and his colleagues also recently came out with $XCMS^2$ for MS/MS data.

For researchers blending NMR and mass spec data, Lindon and his colleagues have also been working on a tool that bridges the NMR-mass spectrometry divide. Their statistical heterospectroscopy, or SHY, works to put NMR and UPLC/MS data from the same samples together by analyzing signal intensities from the molecules as detected by the different methods. "You get a bit of information from the mass spec and a bit of information from the NMR, you can put the two together to identify molecules," says Lindon.

Databases

With the molecules in hand, the identity of the metabolites can begin to be uncovered, though it isn't always possible when they don't correspond to a known metabolite. "There are still a lot of unknowns in terms of compounds that people see or identify. If you were to take a sample from a person or a plant and use our standard libraries of known endogenous metabolites, you still won't be able to identify all the compounds, or all the peaks," Wishart says.

A few database projects — including efforts by Wishart and Siuzdak — are attempting to index and curate all the known metabolites. "Unlike in proteomics or genomics where we can say we know all the amino acids and all the bases and therefore the library or the alphabet is known, the alphabet isn't really fully known for all the things that we expose ourselves to," Wishart adds.

"We can use that statistical correlation to prove those two peaks are linked across hundreds or even thousands of samples."

Starting in January of 2005, Genome Canada funded the Human Metabolome Project; part of its mandate was to catalogue and consolidate all naturally occurring metabolites. It contains about 2,500 metabolites, culled from the literature and confirmed with NMR, LC/MS, or GC/MS, as well as from the group's own experimental data.

Siuzdak and his colleagues are working on Metlin, a depository for mass spectral metabolite data. It currently contains about 23,000 molecules, and Siuzdak says they are adding more to it constantly. The 1966 paper, says Siuzdak, said the main problem with using GC/MS was that there are so many molecules that are unknown and there's no comprehensive database. "What happens since then is now there's a database that has well over 10,000 molecules in it," Siuzdak says.

While these projects and others,

such as Riken's SpinAssign and the Madison Metabolomics Consortium Database, have made progress in cataloguing metabolites, estimates place the number of metabolites in the tens of thousands. The databases have a long way to go before they can be considered anything close to exhaustive. "[Metabolomic databases] still have a ways to go. They are not as robust as Blast or Mascot," Wishart says.

Not alone

Metabolomics isn't the be-all and end-all. Once the data is gathered and analyzed, with the metabolites identified, metabolomics often leads to new questions that can be followed up by using the other arms of systems biology. Because metabolomics may be more reflective of phenotype, as Siuzdak says, using it in combination with "the genetic information that we have, it gives us a really interesting story."

Andrew Lane agrees. "You can't do just one of the 'omics on its own," he says. "Once you've found something out from a metabolic pathway, you need to go back and verify that, OK, we're positing that this metabolic pathway has increased activity, that implies that there's either increased gene expression for those enzymes in that pathway or that some of the enzymes in that pathway have become more active by post-translational modification or by allosteric regulation. You have to look at gene expression, protein level, and protein post-translational modifications."

But that integration across the field is a challenge, not only for metabolomics, but for systems biology as a whole. "There's absolutely no point in just concentrating on one 'omics. We have to be able to integrate data across all the 'omicses," Lindon says. "Making sense of data that we collect at the different levels of the 'omics — genomics, transcriptomics, proteomics, and metabonomics — understanding all of that in the context of systems biology is very, very important. It's where we're going."

Can you afford to lose on content?

With Tecan's new PowerScanner you're ready for analyzing ultra high density microarrays. Integrated ozone filtering, auto-processing and internal calibration provide perfect quality conditions. Our premium microarray scanner gives you 2 µm pixel resolution at benchmarking sensitivity and accuracy. Compatibility of data formats and automated batch analysis is provided.

This system is available soon.

- 2 – 40 µm resolution
- < 0.04 fl/µm² detection limit
- < 2.5% CV uniformity
- 10⁵ dynamic range

More Reasons to Talk to Tecan

www.tecan.com/PowerScanner USA +1 919 361 5200 ext. 549 Europe +49 79 51 94 170 Asia +81 44 556 7311

Liquid Handling & Robotics | Detection | Sample Management | Components | Service & Consumables

© 2008 Tecan Trading AG, Switzerland, all rights reserved.

ROSETTA SYLLEGO™

Bring it all together with the Rosetta Syllego™ system.

The Syllego data management and analysis system for genomics studies connects relevant genomics, gene expression, and phenotypic study data, and also connects you with collaborators, public resources, service providers, and functional groups. Take advantage of the Syllego system's unified data management solution and multiple workflows for CNV, eQTL, and GWA studies.

Visit us at ASHG to learn more. Attend our Syllego system workshop on November 12, 2008 or stop by our booth #616.
For more details, visit **www.rosettabio.com/syllego/ashg.**

1.800.737.6583 (U.S.)
+1.206.926.1220 (Outside the U.S.)
sales@rosettabio.com
www.rosettabio.com/syllego

Copyright © 2008 Rosetta Inpharmatics LLC. All rights reserved. Syllego is a trademark of Rosetta Inpharmatics LLC.
The Rosetta Biosoftware convergence icon is a registered trademark of Rosetta Inpharmatics LLC.

STRUCTURAL VARIATION

Counting on Copy Number

Research into copy number variation is highlighting just how complex genomic differences are. A glimpse at the latest findings and approaches to understanding CNV regions and what they mean.

BY MEREDITH SALISBURY

Copy number variation broke onto the scene four years ago, when what was once thought to be an occasional genomic glitch turned out to be an incredibly common mechanism of DNA variation.

Charles Lee, an associate professor in the pathology department at Harvard Medical School, worked on one of the earliest studies revealing this phenomenon. A research project involving 39 healthy control patients showed a significant number of "gains and losses which we weren't expecting to find," Lee says. "We thought it was artifacts, [but] when we started validating them, they weren't artifacts. They were true gains and losses."

The real surprise was how much of an impact these variant regions were having. "What we did not anticipate is that copy number variations are so abundant that they affect [a] greater number of bases than single nucleotide polymorphisms do," says Victor Guryev, a member of Edwin Cuppen's lab at the Netherlands Institute for Developmental Biology.

CHARLES LEE

Today, scientists posit that variation in copy number has even more of a role in disease association than SNPs do.

Basic biology

Much of the research into CNVs in these early days is simply geared toward understanding what these elements are doing, how they came to be, and how it's possible for organisms to have such different copy numbers and still appear the same.

Matthias Platzer at the Leibniz Institute for Age Research and Fritz Lipmann Institute focuses on the 350 KB region at 8p23.1 in the human genome, a cluster of well-researched defensin genes that have extraordinary range in copy number variation. "The range of variation is really huge, from two copies as a normal diploid genome to up to 12 to 14 copies of that entire region," he says. The question is, how can people vary by so many bases of sequence and still show no difference in phenotype?

An answer may lie in understanding how copy number variation happens, but that's elusive at this point, says Guryev. "Finding the molecular mechanisms responsible for CNV formation remains [a] challenge. Change in copy number is caused in various ways, such as non-allelic homologous recombination, non-homologous end joining, instability of tandem repeats, or transposition of mobile elements," he notes. "We still do not have a complete overview on how these mechanisms contribute to the diversity of structural genome alterations."

Some light has been shed on the issue by research such as Noah Rosenberg's, which focuses on CNV in human populations. Rosenberg, an assistant professor at the University of Michigan, has been studying populations globally to determine patterns of variation. "For the most part, the pattern of copy number variation in worldwide populations matches what we expect in terms of SNPs and microsatellites," he says. "That's telling us the history of copy number variants largely matches the human history as a whole."

Rosenberg's work also demonstrated evidence of natural selection at work on these variants. "We noticed that many of the copy number variants were rare," he says, indicating that "there's some negative selection operating against at least a reasonable fraction of these variants.

Otherwise, some of these would be a bit more common."

Another piece of the puzzle was supplied by Harvard's Lee, whose involvement with the Structural Genomic Variation Consortium — a partnership with Harvard, Sanger, and Toronto's Hospital for Sick Children — led to research into population differences of copy number in amylase genes, which are involved in starch digestion. As hypothesized, populations with higher levels of starch in their diets had higher numbers of the gene. "That was the first time one of these CNV regions was shown to be under positive selection," Lee says.

Move to models

Of course, biologists are taking advantage of their go-to resource for better understanding bizarre events in the human genome: CNV research into model organisms is taking off. Lee says that the turn to animal models lagged; when he began to study variation in chimps and macaques, "I got a sense … that there was clearly a deficiency of work being done in other animals," he says. A recent paper from his team showed a first-pass look at these primates, demonstrating that copy number variation does indeed affect their genomes. As it turns out, Lee says, "when you have segmental duplication in the genomes of organisms, they do have the ability to foster the creation of copy number variants."

Following that, Lee's group has been delving into zebrafish, which also has segmental duplication. His results aren't yet published, but Lee believes the level of variation he's seen in zebrafish will be "a very eye-opening experience" for the research community. So much variation could play a significant role in the run-of-the-mill genetic experiments done on these organisms and will have to be controlled for, he adds.

Edwin Cuppen's group is using inbred rat strains to try to get a purer view of copy number variation. Evaluations of these regions in rat are still at a low-resolution phase, Cuppen says, but he believes the method of using inbred strains will remove a lot of the background noise that can't be controlled in most organisms. So far, he says, it's clear that copy number changes are responsible for "quite a few expression differences" in the organism.

Complex techs

Cuppen notes that copy number events are more complex than initially suspected. Copies don't appear faithfully and in whole; they can be duplications combined with inversions and small deletions, for instance, making them much more difficult to detect comprehensively. Because of that, Cuppen uses a number of technologies to study these elements, and says that just one platform isn't enough to track this kind of variation. His group uses standard array CGH technology with paired-end sequencing as well as optical mapping for the rat studies.

"A promising new technology for detecting structural variants is combination of paired-end mapping and next-gen sequencing," Guryev says. "However, it will require even higher sequencing throughput and price reduction before we can use it for such applications as diagnostics or association studies."

One challenge is that standard technology — generally speaking, array CGH — isn't precise enough to quantify copy number variation, says Liebniz's Platzer. "In these techniques you see just that there are more than two, or maybe four or five, and you have no information about the exact copy number," he adds. His team worked with MRC Holland to develop MLPA, or multiplex ligation-dependent probe amplification, a technology specifically designed for the defensin gene region he studies. MLPA increases probe density to get a high-res view of the region, and Platzer says that "from our point of view, this is the most quantitative approach at the moment."

Rosenberg says there's still a need for better quality control, especially to reduce false positives, and that technology needs to evolve to account for more complexity. His population study looked for five states of variation — homozygous or heterozygous deletion, normal, and homozygous or heterozygous addition — and he says current tools make it difficult to go beyond those states.

Link to disease

Whether technologies improve or sequencing gets cheap enough to enable whole-genome scans for copy number variation, the ultimate aim is the same: figuring out how these changes contribute to disease. Research has already shown that CNVs are tied to schizophrenia and autism, among other diseases, and scientists expect that trend to pick up steam. This "highlights the importance of copy number changes in disease etiology," Guryev says.

In this sense, CNVs are "like risk factors," says Lee. They may eventually help stratify patients to show which will respond better to one drug than another, for instance.

Lee believes that a connection to cancer is imminent. "I think we're going to find that there are some CNVs" that increase predisposition to cancer, he says. "I think it's coming right around the corner."

The broader your view,
the more associations you'll find.

Get the solutions that bring SNPs and CNPs together for the first time.

Affymetrix genotyping products allow you to study more variation, providing you with the most complete view of the genome. By combining SNPs and copy number polymorphisms (CNPs), the Affymetrix platform not only gives you a competitive advantage, but also helps advance our understanding of many common diseases. With open software and instrument platforms, and a unique view of SNPs and CNPs, you'll make more relevant discoveries.

To learn more about our genotyping solutions for association studies, visit www.affymetrix.com

©2008 Affymetrix, Inc.

ALTERNATIVE SPLICING

A Complexity that Goes Beyond Genes

Most genes are subject to alternative splicing, but it's still early days in understanding the phenomenon across pathways or on a genome-wide scale. A look at some pioneers in the field and the technologies they've commandeered to make sense of it.

BY JEANENE SWANSON

DIANE LIPSCOMBE AND MICHAEL WOLFE

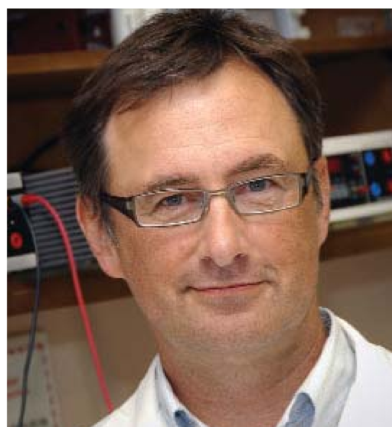
PHOTO: SAM RILEY

Genome Technology Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

For most of the past 10 years, Michael Wolfe has been studying the biochemistry of gamma-secretase, painstakingly detailing its structure, function, and mechanism of action. The main culprits of Alzheimer's disease are short amyloid-β peptides that build up and clump together in the brain. Gamma-secretase works alongside beta-secretase to cleave the amyloid precursor protein into its fatal, truncated cousin. Several years ago, Wolfe began characterizing beta-secretase, but this time with a new tack: determining how alternative RNA splicing affects the enzyme's function. Recently, Wolfe published work that identified alternative splicing events in beta-secretase.

"It's starting to look like most genes in the human genome are alternatively spliced," says Wolfe. A single gene can generate many types of proteins, which lends the relatively small human genome its extreme diversity of expression. That beta-secretase undergoes splicing is not surprising. "It seems to be a key regulatory event, and a way for the cell to control what kinds of proteins are produced from a single gene," Wolfe says.

Alternative splicing has now been found to exist for at least 70 percent of genes. After DNA is transcribed into an mRNA precursor called pre-mRNA, splicing machinery known as the spliceosome steps in. The spliceosome snips out the introns and patches up the strand to form a finished mRNA, which is then translated into a protein. In alternative RNA splicing events, certain exons are skipped, or specific introns left in, so that the final mRNA sequence can vary, producing different splice isoforms of the same protein.



"It's very anecdotal. People are working on their gene of interest, and they find different isoforms by cloning the genes and [then] try to figure out what this other isoform is doing."

Benoit Chabot
UNIVERSITY OF SHERBROOKE

While there's a lot of interesting pathway analysis being done, the tools to look at this kind of variation are, for the most part, in early stages. Most researchers still use RT-PCR to identify new splice variants or to confirm splice microarray results. Biologists have only just begun to scratch the surface of associating different isoforms with unique functions inside the cell. "Looking at diversity of function is still in its infancy. It's very anecdotal," says Benoit Chabot at Quebec's University of Sherbrooke. "People are working on their gene of interest, and they find different isoforms by cloning the genes and [then] try to figure out what this other isoform is doing. Nobody is doing it in a systematic manner right now." For the most part, large-scale tools like splice arrays and up-and-coming tools such as multiplexed PCR and high-throughput sequencing are just beginning to enter the alternative splicing research world.

Protein by protein

Like Wolfe, Brown University's Diane Lipscombe uses splicing as a window into the expression diversity of her protein of interest, voltage-gated calcium channels. Her lab was one of the first to clone these proteins in neurons, where they're important in modulating processes as diverse as gene transcription and neurotransmitter release, and have been linked to epilepsy and migraines. In her current studies, Lipscombe looks at how neuronal-specific factors affect splice variation in the channel and end up fine-tuning its structure and behavior. She typically uses gene databases and comparative sequence analysis to identify a hit, and then goes back in with PCR to see how the splice isoform may be differentially expressed in the tissue of interest. "It's very old-fashioned," she says.

In work published in 2004, Lipscombe found that a particular isoform of the channel was enriched in nociceptors, neurons that can sense and signal pain. She noted in a recent study that this channel is not only more sensitive to neurotransmitters, but it's also more sensitive to opiates like morphine. "We've long thought that the expression of different splice isoforms probably underlies a lot of the differential effects of drugs in different pathways," she says.

Alternative RNA splicing can vary depending on tissue and stage of development, among other things, and in most circumstances there is a healthy balance between differen-

tially expressed isoforms. Existing side by side in certain percentages, variant transcripts can regulate gene expression by turning protein manufacture on or off. When aberrant splicing occurs — for instance, in some cancers — pathways commit signaling errors which lead to downstream trouble. Biologists are looking for not just whether a particular isoform is there, but whether it's an actual splice variant or just an error in transcription, whether it's tissue-specific, and whether it occurs at a level that's meaningful enough to have any noticeable effect.

Peering into pathways

Scott Friedman, chief of the liver disease division at Mount Sinai School of Medicine, knows about devoting a lot of time to a single gene. For more than a decade, he's been studying alternative splicing in a tumor suppressor gene called Krüppel-like factor 6 and how a splice isoform, KLF6 SV1, affects liver cancer. He cloned the gene for KLF6 10 years ago, and after his research turned toward alternative splicing, found that SV1 antagonizes full-length KLF6 suppressor activity. "Transient splicing can be a very subtle fine-tuning mechanism for adjusting cell function," he says, "whereas in cancer, it looks like it's kind of a constitutively on switch that affects cells."

> "Once you recognize it's another level of regulation and you look for it, it's amazing how prevalent these kinds of regulatory pathways are."

SCOTT FRIEDMAN

What they found is SV1, "which is really composed of three of the four exons of the tumor suppressor, seems to be drastically up-regulated in many late-stage cancers," says co-author and fellow Mount Sinai scientist John Martignetti. He adds that there seems to be an interaction between the two variants depending on how much of each is present. Friedman and Martignetti recently collaborated on work that fleshes out some of SV1's mechanism of action; using RNAi, they were able to show that knocking down the Ras signaling pathway could decrease SV1 production and inhibit tumor growth. Martignetti's work also looks at prostate and ovarian tumors, and what's going on in signaling dysregulation in these cancers. He's seen up-regulation of SV1 in prostate cancer, where it appears to play a role in many different pathways. "It seems to be involved in changes in proliferation, in metastasis, and even in angiogenesis," he says.

To detect and measure different splice variants, they also use PCR and are trying to develop splice-specific monoclonal antibodies. RNAi using siRNA has been a very effective tool, they say, in that it can be specific enough to distinguish between the tumor suppressor and its oncogenic variant.

Brown's Lipscombe, however, says the hardest part isn't identifying the variant at the RNA level. "Where we are incredibly limited, where we have a huge hurdle to overcome, is to be able to distinguish at a protein level," she says. "It's a problem. You don't have that much control." As an example, there is a calcium channel variant differing by an exon that encodes only two amino acids, but Lipscombe hasn't yet been able to create antibodies for the two nearly identical isoforms. As a stopgap solution — it's harder, it's more expensive, but it gets the job done — she's created mice that express one or the other.

Despite how difficult it may be, picking up the subtle differences between isoforms is one of the most important steps in pathway analysis. "By having recognized the importance of splicing in [the KLF6] gene, it greatly sensitized our interest in looking for splicing as an explanation for other biologies, using different genes," Friedman says. "Once you recognize it's another level of regulation and you look for it, it's amazing how prevalent these kinds of regulatory pathways are."

Microarrays to market

While tool development is still underway for this field, genome-wide analysis is definitely in swing. In fact, several vendors are hard at work co-opting gene expression arrays to study alternative splicing. Affy's Exon Array has become popular for genome-wide expression analysis, and while ExonHit Therapeutics and Jivan Biologics offer both exon and exon splice junction arrays, the latter have proven to be the truly useful tool in the toolbox.

Jivan's Jonathan Bingham says that while arrays that probe exon bodies give you a lot of information, "you're left with the problem, how do those pieces fit together? And you get that answer more directly if you have probes for the splice junctions."

Because most genes are alternatively spliced, says ExonHit's John Jaskowiak, "you need to start taking a higher degree of interrogation of the genome." ExonHit has been around for 10 years, and the company offers array products for both basic research as well as for therapeutic and diagnostic research. It runs an internal research program for Alzheimer's disease, measuring changes across many samples and tissues. "You've got this capability of multiplexing on an array where you don't have as much flexibility to do that much coverage [with] RT-PCR," Jaskowiak says.

To create its arrays, ExonHit mined EST databases and other cDNA genomic information, and then partnered with Affy and Agilent to print them. They offer genome-wide arrays for human and mouse, as well as other arrays specific to druggable-target gene families, such as apoptosis, cytokines, or kinases.

Jivan also offers genome-wide arrays for human, mouse, rat, fly, and more, as well as targeted arrays for specific applications like oncology or toxicology. "Most of our focus has been on microarrays until more recently," says Bingham, who notes that because data analysis poses such a problem, the company will be making refinements to its software in the future. Most of their initial customers have been academic researchers, but Bingham says he's seen a shift toward using the arrays for diagnostic purposes like studying biomarker signatures, drug response, or toxicity. And with the ability to look across many genes at once with arrays, it's easier to study splicing regulation — a separate field that looks at hundreds of interacting genes and accessory proteins that modulate how and when a gene is spliced.

Not perfect yet

One problem with using microarrays, of course, is that they can't be used to find new variants since only known variants are spotted

BUILDING BLOCKS TO BETTER MICROARRAY HYBRIDIZATION

Biomicro® Systems now offers the MAUI® Product Suite to boost the accuracy, consistency, productivity and cost-effectiveness of today's microarray research.

MAUI HYBRIDIZATION SYSTEM.
BETTER HYBRIDIZATION WITH ACTIVE MIXING™!

No Mixing | MAUI - Active Mixing

Data Courtesy of:
Patrick O. Brown Laboratory, Stanford University School of Medicine

APPLICATIONS

- Gene Expression
- array CGH
- miRNA
- Protein
- FISH
- FFPE . . . and more!

Schedule a DEMO today to see how the MAUI Product Suite can boost your microarray research!

biomicro.com | 1.800.454.1485

MAUI Hybridization System (4 and 12-Bay Models)

Over 800 installations worldwide.

MAUI Mixer Hybridization Chambers

Multi-chamber formats (10 - 40 µl).

MAUI Wash System

First we made microarray hybridization better. Now, we make slide washing easier!

Spot-On Solutions™ for Microarray Research

TM & © 2008 BioMicro Systems Inc. All rights reserved.

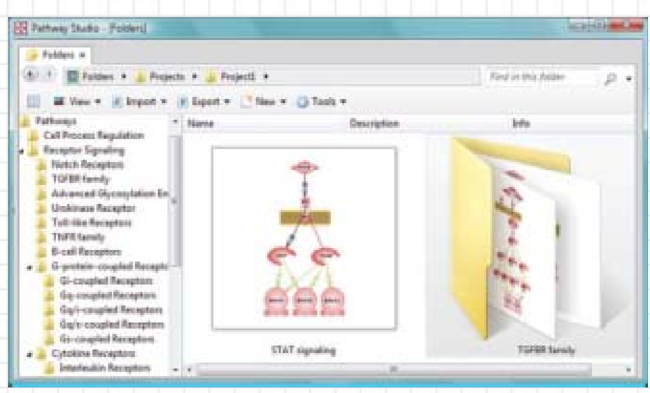
PATHWAY STUDIO®

Visualizing the Science

Rapid, Real-Time Biological Knowledge Mapping

Pathway Studio streamlines hypothesis generation by creating colorful visual networks, and helps find those pesky "needles in the haystack" fast and efficiently. Powered by MedScan technology, it pulls together, organizes and integrates knowledge, not just stand alone findings, from many data sources.
Just minutes a day gets you this morning's most recent published findings in your field of expertise: RNAi, target discovery, protein or biomarker research to name a few. It all starts with your PubMed search and ends with a mechanistic model compiled by you and focused for your research program. Always Refreshed, Always Relevant, Always Informed.

REGISTER FOR A FREE TRIAL AT:
www.pathwayanalysis.com



Visit us at RNAi Europe, Stockholm, Sweden
16-18 September 2008

Ariadne Genomics, Inc. | 9430 Key West Avenue Rockville, MD 20850 | 240.453.6296 MAIN 866.340.5040 TOLL FREE | info@ariadnegenomics.com www.ariadnegenomics.com

down. Even if the chip is meant to cover the entire genome, not every splice variant is known for most genes, nor to what extent minor variants play a role in gene regulation. For disease genes, which tend to produce splice variants in precise percentages, finding minor variants is especially important.

"In terms of looking at splicing, we're fairly low tech right now," says Lipscombe at Brown. "What we really need to know is whether a particular mRNA that encodes for a given isoform is present in sufficient quantity that we can say, 'Oh, yes, this must be meaningful.' For sure we could use microarray," she says, but correlating function usually comes back to isolating a variant from one particular population of cells. "Microarray analyses are probably better suited to scanning a whole bunch of different tissues and looking for differential distribution patterns, and so for us, because we know what we're going for — we're just looking at one particular gene — the arrays are not necessarily better than what we're doing right now."

Most of the microarrays are based on EST database information, so they tend to find large exons, Lipscombe adds, ignoring possibly critical drug target regions because they vary by only a few amino acids. "I think there is an increase in databases but there's still a lot of information missing. Right now it's not refined" enough to discover tiny variations or variants in small or unique populations of cells, she says.

Jernej Ule of the MRC Laboratory of Molecular Biology in Cambridge, UK, says he thinks most researchers use exon arrays since it would be too much bioinformatics work to design an array that probes exon junctions. But in coming years, "splice junction microarrays should become standard to replace the current microarrays to characterize disease-related changes and possibly even for diagnostics," he says.

LUIZ PENALVA

Another limitation to using arrays is that even when a variant is identified, it has to be validated with PCR anyway and tested with functional assays. "I think, in general, the technology is very, very good — and what is important now is to apply it and understand how these changes that are detected are actually being regulated in the cell and which ones are relevant," Ule says. Determining which variants are actually being selected for and which are just errors is an even bigger challenge to the splice research community at large.

> **Determining which variants are actually being selected for and which are just errors is an even bigger challenge to the splice research community at large.**

A significant bottleneck for arrays is data analysis. Luiz Penalva, assistant professor at Children's Cancer Research Institute at the University of Texas Health Science Center in San Antonio, says performing the analysis of his experiments, which attempt to identify splice variants in glioblastoma, was the hardest part. "This is one of the major problems with alternative splicing microarrays nowadays," he says. While arrays do come with software analysis packages, none of them is perfect; Penalva says his team, which includes a bioinformaticist, usually has to try many different methods to get trustworthy data.

"One thing that we observe when you get array results, it looks like there are some probes there that are simply not giving you any data, or data that doesn't look correct," Penalva says. "Sometimes … it's better simply to discard this data."

Doug Black, a Howard Hughes investigator at the University of California, Los Angeles, uses a full arsenal of tools to study how splicing is regulated and the role these regulators play in neuronal cell differentiation. In mature neurons, he examines how calcium signaling pathways and chronic depolarization can change splicing. While he incorporates both exon arrays and splice junction arrays into his day-to-day work, he says there are tradeoffs to each. With exon arrays, "[it] is a little bit of a pain and you end up having to validate a lot of what you find through other means," he says. However, when using exon junction arrays, plenty of work has to go into probe design, and coming across new variants isn't possible the way it is with a genome-wide exon array. Still, says Black, the splice junction arrays are more sensitive and more reliable.

As a complementary tool, high-throughput sequencing holds much promise. Black is already using a Solexa sequencer to profile splicing under different conditions. He believes that affordable next-gen sequencing will be able to take RNA splice variant detection to the next level. "It's potentially much more accurate and much easier in the analysis," Black says. In fact, a flurry of studies published recently used RNA sequencing, or RNA-seq, to survey the complete transcriptomes of mice, *Arabidopsis*, yeast, and human cells. "In practical terms, it's not there yet," Black adds. "You don't get enough individual short reads to sample all the exons that you want to sample. You don't have enough sequence depth without hundreds of thousands of dollars to actually measure splice variants."

> **"You don't have enough sequence depth without hundreds of thousands of dollars to actually measure splice variants."**

And in the event that you're not wading in funding, Sherbrooke's Benoit Chabot has come up with an affordable alternative to next-gen sequencing. As multiplexed PCR takes off, Benoit has taken advantage of the capacity of newer machines to skip microarrays altogether. Currently he can run 3,000 RT-PCR reactions per day, and he hopes to increase that 10-fold during the next few years. "We're not going to do microarrays," says Chabot. "That means we're not going global as much, but we're going to automate it, make RT-PCR a little bit more high-throughput than what people are using." The approach allows him to look at select splicing events across many tissues and many conditions — experiments that probably wouldn't be affordable using sequencing or microarrays. Whether PCR is used to validate microarrays or as a "replacement for cases where customers are looking at a particular pathway or a particular gene family," says Bingham at Jivan, it will continue to offer up more opportunities in a multiplex setting.

Toward the clinic

As basic research moves ahead with identifying new variants and validating their functions in the cell, clinical research and drug development are already looking closely at how splice variants are dysregulated in disease. Often variants will be expressed alongside one another, but in a healthy percentage. When alternative splicing goes awry, one variant may be more or less expressed, leading to signaling imbalances and disease. It's no surprise that finding drug targets would incorporate the study of splicing.

"Initially, the most interest was coming from the academic world, but what we've seen more recently is that drug companies are starting to do pretty decent-sized studies," says Jivan's Bingham, "because splicing offers potentially more information for biomarkers than gene arrays alone." Splice arrays are used to look at expression signatures for disease state or disease progression, to screen for specific isoform drug targets, or to mark drug response "where splicing changes after a drug is administered," he adds. So far, he notes, the most interest has been in arrays for cell surface and toxicology genes.

In its work developing tools to study Alzheimer's disease, ExonHit began using splicing arrays as a diagnostic to screen patients for clinical trials. By measuring RNA in circulating blood, scientists are able to determine whether someone has Alzheimer's, another form of dementia, or a different disease altogether. Today the only way to accurately pinpoint the disease is post-mortem, so splice signatures have great potential in clinical Alzheimer's research. "The interesting thing about that signature is there are a lot of splicing isoforms that are present," says Jaskowiak at ExonHit. "In many cases, it's the ratio of isoforms that is important."

Harvard's Wolfe hopes that identifying splice variants that cause Alzheimer's will eventually be useful in finding effective drug targets. Today, methods for slowing down the progression of the disease mostly revolve around tweaking gamma-secretase production, but gamma-secretase also plays a central role in the highly conserved Notch signaling pathway. Wolfe's recent work found that beta-secretase also undergoes alternative splicing and could be an alternate drug target. By interfering with its alternate splicing events, one could interfere with beta-secretase function, he says. "Alternative splice isoforms exist. They can vary depending on the cell type, and if we shunt splicing down these alternative pathways, we can lower amyloid production."

Wolfe has also been studying the impact of splicing on tau, a gene associated with a related, non-Alzheimer's form of dementia. He's found that about half the known mutations in the tau gene change its splicing "and somehow that

leads to the self association of tau in clogging up neurons," he says. "What's regulating the alternative splicing of tau? Can we pharmacologically step in and tweak the system?"

Discovering drugs that target one point in a splicing pathway can be difficult considering the number of variables. A recent study out of Chabot's Sherbrooke lab looked at how anticancer drugs affect splicing of Bcl-x to promote apoptosis. "It's known that many types of apoptotic genes are alternatively spliced to produce pro-apoptotic variants or anti-apoptotic variants," Chabot says. To his surprise, he says, no one had ever systematically looked to see if pro-apoptotic drugs actually initiated the apoptotic pathway. In his study, he used 20 drugs on five different cancer cell lines to see how they affected apoptosis. While all the drugs shifted splicing of Bcl-x in the right direction, "it does not do it systematically in all cell lines. Some cell lines respond to it; other cell lines don't, depending on the drug." For other alternatively spliced apoptotic genes, this wasn't the case — for some drugs it went in the right direction and for others it didn't. "It's very complex, and we cannot assume that taking an anticancer drug will always go in the right direction," Chabot adds.

There are a number of ways that alternative splicing events could be manipulated, all of which one day might be applied in the clinic. One approach is to use antisense molecules that would bind to disease-causing transcripts, effectively turning them off. Another might be to make use of nonsense stop codons, which in normal alternative splicing events tell the translation machinery to stop before the full-length protein is complete. In the case of tau, the mRNA forms a tiny loop structure that Wolfe thinks could somehow be made to bind to a small molecule to steer splicing toward one isoform or another. "If there's a structure, it probably has pockets where small molecules can bind, so to the degree we can find structure in the message that regulates splicing, we might identify therapeutic targets to get very specific effects," he says. "There are some attempts at therapeutics, but it's pretty early days."

More Resolution

New Instruments!

MDS Analytical Technologies introduces the new Axon GenePix® 4300A and GenePix® 4400A microarray scanning and analysis systems, which combine increased imaging resolution with a host of enhancements to offer the highest performance of any slide-based microarray scanners. Our open platforms will enable image acquisition of any fluorescent microarray, even the newest ultra-high density formats.

Axon GenePix 4300A and GenePix 4400A Scanners

- Increased resolution: 2.5 µm per-pixel imaging in the GenePix 4400A system for the highest-density arrays; 5 µm resolution in the GenePix 4300A system can be upgraded after purchase
- Controlled uniformity: obtain more consistently reproducible results
- Improved sensitivity: revised optics give unmatched limit of detection
- Superior configurability: choose from four lasers and employ up to 16 different emission filters
- Enhanced automation: minimize your interaction and your time with a fully automated system

Axon GenePix Pro 7 Software

- Powerful multiplexing: independent image acquisition and analysis of multiple arrays per slide
- Complete hardware compatibility: control any Axon GenePix scanner, new or old
- Operating system flexibility: use 32- or 64-bit versions of Microsoft Windows Vista or Microsoft Windows XP

tel. +1-800-635-5577 | www.moleculardevices.com

i switched!

"The sensitivity of Illumina technology gives me a real edge in transcriptome sequencing. I'm seeing more, and seeing it faster, than ever before. I'm definitely going to continue sequencing mRNAs. I'm not going back to exon or tiling arrays."

Rickard Sandberg, Ph.D.
Group Leader, Department of Cell and Molecular Biology
Karolinska Institutet, Stockholm, Sweden

See the transcriptome like you've never seen it before. Measure alternative isoforms. Unveil novel structures. Discover coding SNPs. mRNA-Seq for the Illumina Genome Analyzer takes your research further than you ever thought possible.

A different kind of gene expression company.

www.switchtoi.com?gt

GENE EXPRESSION
SEQUENCING
GENOTYPING

SEQUENCING | Upstream

INSEQUENCE

Barcode of Life Project Leans on Sequencing

An international consortium of researchers says it plans to increase by more than 10-fold the catalog of eukaryotic species that are tagged by a DNA barcode, and to develop new barcoding technology to identify specimens rapidly and inexpensively.

The first phase of the project, which will generate a library of barcoded species, will largely involve Sanger sequencing technology, according to one of the organizers. Second-generation sequencing technologies will find applications in environmental barcoding studies later on, and a long-term goal of the project is to develop a handheld barcoding sequencer.

The initiative, called International Barcode of Life Project, or iBOL, currently involves 26 countries. Planning for iBOL started last year at a workshop at the University of Guelph in Canada that brought together a variety of international researchers with a shared interest in barcoding.

Barcoding involves sequencing a short, standardized gene region that differs between species. In animals, for example, researchers use a portion of the mitochondrial cytochrome c oxidase I gene as a barcode.

iBOL's first aim is to create a reference library by barcoding 5 million specimens representing 500,000 species within five years. This project, set to begin next year, will significantly expand the current library, which comprises approximately 41,000 barcoded species.

iBOL is currently focused on raising at least $100 million of its $155 million budget from various funding sources around the world, according to a project outline published by the consortium in July.

What distinguishes this project most from other large-scale efforts is that each sequence read derives from a different sample, according to Paul Hebert, director of the Biodiversity Institute of Ontario at the University of Guelph and an iBOL organizer.

To generate a 650-base read for each of hundreds of thousands of samples, "Sanger sequencing technology is really the only feasible way to go," he says.

— Julia Karow

DATAPOINT

2

HELICOS BIOSCIENCES ANNOUNCED RECEIVING A SECOND ORDER FOR ITS INSTRUMENT, BUT DID NOT NAME THE RESEARCH CENTER PLACING THE ORDER.

Sequencing Notes

OXFORD NANOPORE TECHNOLOGIES acquired exclusive rights to develop and market nanopore technologies from **HARVARD UNIVERSITY** and collaborators at the **UNIVERSITY OF CALIFORNIA, SANTA CRUZ**, and the **NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY**. The company says label-free nanopore technology could help to reduce the cost of DNA sequencing.

In its second-quarter earnings report, **HELICOS BIOSCIENCES** said its grant revenue was $251,000 for the period, while its net loss grew to $11.9 million, an increase of 47 percent from the same quarter in the previous year. President **STEVE LOMBARDI** said problems with reagent stability hindered efforts to get new sales. The company reported no revenue from product sales during the quarter.

FUNDED GRANTS

$577,448/FY 2008

MOLECULAR ENGINEERING APPROACH TO STUDY LONG-TERM SYNAPTIC PLASTICITY

Grantee: Jingyue Ju, Columbia University
Began: Feb. 1, 2008; Ends: Jan. 31, 2012

Ju and his team will continue development work on their Massive Parallel DNA Sequencing Chip System for use in digital gene expression, with the aim of doing large-scale expression studies in single neurons. According to the abstract, the tools will be tested on the memory-forming network of *Aplysia*, a unique model organism for neurobiology.

$236,530/FY 2008

MICROBIAL COMMUNITY PROFILING OF SEWAGE CONTAMINATION IN THE GREAT LAKES

Grantee: Sandra McLellan, University of Wisconsin
Began: Jun. 15, 2008; Ends: May 31, 2010

McLellan and her team will use massively parallel DNA sequencing strategies to study the microbial communities and other sewage contamination present in the Great Lakes in the northern US. The lakes serve as drinking water to 40 million people, and bacterial content of the water is unknown.

discovery2diagnostics
CONFERENCE & EXHIBITION

The ONLY scientific event in 2008 that brings together discovery technology, biomarkers, molecular diagnostics and molecular theranostics

October 20-22, 2008
Manchester Grand Hyatt San Diego • San Diego, CA

Three Comprehensive and Interrelated Conference Tracks

1 Sequencing and Detection Technologies

- Learn about advances in miRNA prediction, detection, and analysis methods
- Explore Next Generation Sequencing technologies being used in novel applications
- Evaluate various miRNA purification and labeling methods for how they may impact downstream activity and analysis

2 Molecular Diagnostics

- Learn about companion diagnostics, third world diagnostics, commercialization of microarray-based molecular diagnostics for personalized medicine applications, and biomarker and molecular diagnostic reimbursement
- Hear viewpoints and updates on the regulatory environment important to molecular diagnostic development and success in the clinic

3 Biomarkers

- Gain scientific knowledge for development of effective patient stratification schema
- Hear about new research into miRNA, imaging, metabolomic, and proteomic biomarkers and how they inform diagnosis of a range of diseases
- Learn how functional RNAi molecules are being translated from markers into therapeutics

NEW Special Attractions

- Enhanced topical integration across all three conference tracks
- Latest RNAi discoveries and developments impacting biomarkers, molecular diagnostics and molecular therapeutics
- Increased emphasis on maximizing success in the reimbursement of new-to-market biomarkers and molecular diagnostics

PLUS! Three pre-conference workshops addressing critical process and workflow issues

Corporate Sponsor:

Maximize Your Savings by Registering Early
www.Discovery2Diagnostics.com

Organized by

Priority Code: D8176GT

PROTEOMONITOR

Satoris to Launch Alzheimer's Test

While proteomics holds promise as a technology that could potentially help in the early diagnosis of Alzheimer's disease, no one has yet been able to translate proteins into a predictive tool for the debilitating ailment.

But a small molecular diagnostics firm, Satoris, is trying to change that. Its chief executive says the company's protein-based Alzheimer's test may be ready to hit the market by the end of the year.

The company is currently collaborating with the Mayo Clinic to validate results from a study published late last year. The validation work is expected to be completed in September and, depending on the results, it could clear the way for Satoris to hit the market with two Alzheimer's protein panels within the next few months, says Cris McReynolds, president and chief executive officer of Satoris.

Both panels are based on work published in the fall in *Nature Medicine*, in which researchers, including those from Satoris, found 18 proteins associated with Alzheimer's.

In that study, researchers looked at 259 archived blood samples ranging from patients who had no symptoms to those who had advanced Alzheimer's. The resulting 18-protein panel had both sensitivity and specificity of about 90 percent and was able to "pick out the Alzheimer's from a population of dementia and to properly identify those who had AD," according to McReynolds.

Authors of the study also say that the panel was able to identify patients with mild cognitive impairment who eventually were diagnosed with Alzheimer's two to six years later, but McReynolds says it is not clear whether that means the panel is predictive of the disease or that it is only "able to detect the early disease process associated with AD." Though mild cognitive impairment is believed to be a possible precursor to Alzheimer's, some patients with it do not develop Alzheimer's.

— *Tony Fong*

DATAPOINT

$90 BILLION

ANNUAL REVENUE FROM GENERAL ELECTRIC'S TECHNOLOGY INFRASTRUCTURE UNIT, WHICH WILL NOW INCLUDE GE HEALTHCARE.

Proteomics Notes

The Proteomics Research Group of the **ASSOCIATION OF BIOMOLECULAR RESOURCE FACILITIES** is recruiting volunteers to be part of a study looking at different ways to determine quantitative differences in several proteins in six human plasma samples.

MIRACULINS will buy a portfolio of biomarkers from Toronto's **MOUNT SINAI HOSPITAL** to develop a diagnostic for pre-eclampsia. Miraculins recently shifted its focus from proteomics research and development to diagnostics development.

GEORGE MASON UNIVERSITY scientists partnered with **FAIRFAX-NORTHERN VIRGINIA HEMATOLOGY ONCOLOGY** to determine the protein signaling pathways involved in multiple myeloma.

FUNDED GRANTS

$46,826/FY 2008

NOVEL MASS SPECTROMETER FOR COMPREHENSIVE NO-LOSS MS/MS OF ALL STORED IONS

Grantee: Sunnie Myung, Rockefeller University

Began: Jan. 1, 2008; Ends: Dec. 31, 2010

Myung plans to optimize high-capacity ion trap mass spectrometers by isolating the pressure within the ion trap and changing the geometry of the trap to increase resolution. Furthermore, she will couple an orthogonal injection reflectron TOP mass analyzer to the high-capacity ion trap. Then she will use the instrument to study abnormal levels of protein phosphorylation in cancer, diabetes, and rheumatoid arthritis.

$78,500/FY 2008

INTRACELLULAR MYCOBACTERIAL PROTEOME

Grantee: Qingbo Li, University of Illinois at Chicago

Began: Apr. 1, 2008; Ends Mar. 31, 2009

Li plans to investigate the proteome of *M. tuberculosis* H37Rv found within human and murine macrophages using liquid chromatography/linear ion trap-Fourier transform mass spectrometry. Then he will compare those proteomes to identify the active intermediary metabolism pathway, particularly focusing on the half tricarboxylic acid cycles. Li says this may help researchers understand the intracellular persistence of mycobacteria.

© 2008 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries.

The siRNA breakthrough that changes the world of RNAi.

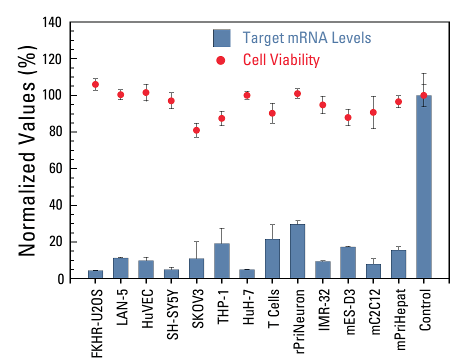
RNAi research is no longer constrained by cell types that are resistant to conventional lipid-based delivery reagents. Thermo Scientific Dharmacon® Accell™ siRNA provides a breakthrough in siRNA delivery. It is specially modified for delivery into any cell type, without the use of a transfection reagent or viral vector, for unprecedented experimental flexibility and discovery.

- Delivery into any cell type, including primary cells
- Free from toxic delivery effects for pure silencing results
- Two-step protocol for extraordinary ease-of-use
- Effective, specific silencing from the RNAi experts

Learn More Today

For more information, call 800-235-9880 or visit us at www.thermo.com/dharmacon

In Europe visit www.thermo.com/perbio



Dharmacon Accell siRNA delivers target knockdown to any cell type

Thirteen difficult-to-transfect cell lines were treated with Accell Cyclophilin B siRNA in delivery media and assayed for mRNA levels and cell viability after 72 hours. Effective silencing and low cytotoxicity was achieved without delivery optimization.

Moving science forward

Thermo SCIENTIFIC

Part of Thermo Fisher Scientific

RNAi **Upstream**

RNAiNEWS

Tacere Evaluates New Drug Opportunities

Having found licensees for its lead expressed RNAi drug, the preclinical hepatitis C therapy TT-033, Tacere Therapeutics continues to seek its next pipeline candidate.

According to Tacere CEO and co-founder Sara Hall, the company is in "evaluation mode" and has been in discussions with a number of undisclosed parties, both from academia and industry, about potentially in-licensing new programs.

"We had some really good meetings at [the Biotechnology Industry Organization's international meeting in San Diego] and we've done some pretty extensive due diligence on a couple things," she says.

Still, the search is early-stage, and Hall declined to comment in detail on the indications or therapeutic areas Tacere is considering.

"Some [possibilities] we are still looking at [and] some we've passed on," says Mike Catelani, Tacere chairman, president, and CFO. "We're not in a position at this point to talk about anything we're looking at, but there are some interesting things out there."

But Hall says the fields Tacere is exploring are hinted at by two recent additions to the company's staff. While both have experience with RNAi, "one has more of a [neurology] background and one has an immunology background," Hall says. Catelani notes that Tacere now has six full-time employees.

Despite Hall's reticence, she notes that Tacere expects to find its next pipeline program outside of its own labs and that it will likely remain focused on expressed RNAi.

"We don't really want to start over again like we did with TT-033 because that's a long, hard road," Hall says. "We'd like to pick up something that's at the translational research stage."

TT-033 was originally developed by Avocel, a company Hall co-founded, which was acquired by Australian expressed RNAi firm Benitec in 2004. Following a sweeping corporate reorganization, Benitec licensed the drug's worldwide rights to Tacere in late 2006.

— *Doug Macron*

DATAPOINT

2009

THE YEAR DURING WHICH SIRNAOMICS PLANS TO BEGIN A PHASE I TRIAL OF ITS MULTI-SIRNA COCKTAIL FOR OCULAR DISEASES

RNAi Notes

ISIS PHARMACEUTICAL is conducting a phase II trial of mipomersen, its antisense-based drug to treat heterozygous familial hypercholesterolemia. **GENZYME** exclusively licensed rights to the drug.

ALNYLAM PHARMACEUTICALS exclusively licensed intellectual property to RNA activation, a gene up-regulation technology. The agreements are with **UNIVERSITY OF TEXAS SOUTHWESTERN MEDICAL CENTER**; the **UNIVERSITY OF CALIFORNIA, SAN FRANCISCO**; and the **SALK INSTITUTE FOR BIOLOGICAL STUDIES**.

PFIZER began a phase II trial of PF-4523655, an siRNA-based drug licensed from **QUARK PHARMACEUTICALS**, in patients with diabetic macular edema. It targets a gene involved in angiogenesis, vascular permeability, and retinal neuron death.

FUNDED GRANTS

$229,600/FY 2008

NOVEL TUMOR SUPPRESSOR GENE DISCOVERY IN PANCREATIC CANCER

Grantee: James Eshleman, Johns Hopkins University
Began: Sep. 1, 2007; Ends Aug. 31, 2009

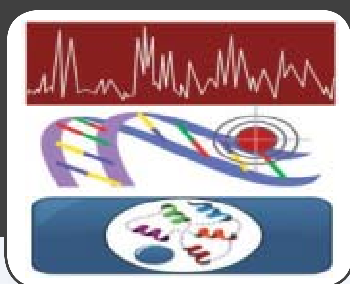
Eshleman will be using RNAi-based techniques to find novel pancreatic cancer tumor suppressor genes. First, he will transduce non-tumorigenic and weakly tumorigenic pancreas cell lines with an RNAi library and grow them up in agar and in nude mice. Then, he will select tumorigenic cell clones and sequence, test, and validate the RNAi.

$258,052/FY 2008

INDUCIBLE RNAI IN SPERMATOGONIAL STEM CELLS

Grantee: Jon Oatley, Pennsylvania State University
Began: Jun. 1, 2008; Ends May 31, 2010

With this grant, Oatley will be seeing if RNAi can lead to advances to treat male infertility. He plans to determine the efficacy of vector-based RNAi on silencing gene expression in spermatogonial stem cells, as well as to evaluate the Tet-On system for inducing RNAi in those cells, and, finally, to determine the efficacy of inducible RNAi to silence essential spermatogonial stem cells' self-renewal genes.



ADVANCES IN METABOLIC PROFILING
EUROPEAN BIOMAKERS SUMMIT
PROTEOMICS EUROPE

16-17 October 2008 Lisbon, Portugal

THREE POPULAR EVENTS COMBINE

Welcome to Advances in Metabolic Profiling, European Biomarkers Summit and Proteomics Europe, taking place in the historic city of Lisbon.

The distinguished faculty includes:

Jörg Baumbach ISAS Institute for Analytical Sciences
Dolores Cahill Conway Institute, University College Dublin
Herbert Fritsche MD Anderson Cancer Centre
Julian Griffin University of Cambridge
Elaine Holmes Imperial College London
Thomas Joos NMI Natural and Medical Sciences Institute, University of Tübingen
Douglas Kell Manchester Centre for Integrative Systems Biology, University of Manchester
Helmut Meyer Medical Proteome Centre, Ruhr University Bochum
Mathias Uhlén AlbaNova University Center, Royal Institute of Technology Stockholm

Corporate Sponsors

For sponsorship and exhibition opportunities contact Paul Raggett
tel: +44 (0) 1787 315117 **email:** paul.raggett@selectbiosciences.com

Agenda Topics

Advances in Metabolic Profiling
- Clinical Applications of Metabolomics
- Fluxomics and Pathway Discovery
- Drug Discovery and Development
- Plant Metabolomics
- Nutrigenomics and Metabolomics
- Microbial Metabolomics
- Novel Technologies for Metabolomics
- Bioinformatics and Data Handling for Metabolomics
- Data Fusion – Bringing Together Different Omic Technologies

European Biomarkers Summit
- Disease Specific Biomarkers
- Biomarkers in Genomics, Metabolomics and Translational Medicine
- Molecular Diagnostics, Personalized Medicine and Pharmacodiagnostics
- Biomarker Assay Development

Proteomics Europe
- Protein Arrays and their Applications
- Protein Expression
- High Throughput Proteomics
- Bioinformatics
- Disease-Related Proteomics

METABOLICPROFILING.COM
BIOMARKERSSUMMIT.COM
PROTEOMICSEUROPE.COM

SELECT BIOSCIENCES
Delivering the Difference

BIOINFORM

Scientists Debate Publishing Future

At the Intelligent Systems for Molecular Biology conference, researchers weighed in on the role that bioinformatics tools will play in the future of scientific publishing.

During the conference, several speakers discussed new text-mining tools and other methods for extracting information from scientific papers. For example, Carnegie Mellon's Robert Murphy spoke about his work on parsing information about images from the biological literature. He and his colleagues have developed SLIF, Subcellular Location Image Finder, a platform that can extract information from captions and figures containing fluorescence microscopy images.

Murphy stated in his presentation that he believes these types of systems will become more widespread but in the meantime it will be important to find ways to "improve practices of defining content," which would make text- and image-mining easier.

One notion he described is the idea of "structured digital caption" that would not show up in the printed paper but would be encoded in the XML file that describes images that are part of a publication. "That makes the parsing of a figure easier," he said.

Yale's Mark Gerstein favors the idea of linking databases and journal articles so that scientists can track a given gene annotation in a database back to the published paper. "There is no good framework for browsing through the genome in the framework of publications," he said. While loading that information into one monolithic database may not be possible, federated queries across structured, ontology-oriented abstracts could help, he suggested.

He said his group "is very keen on this idea of structured abstracts" as a small start to enable a connection between journal articles and databases. Since authors already write the abstracts, this concept would ensure the high quality of the machine-readable abstract. Those texts in turn could be the training set for a more large-scale machine reading project. — *Vivien Marx*

DATAPOINT

$3 MILLION

COMPUGEN'S NET LOSS FOR THE QUARTER ENDING JUNE 30, 2008, INCLUDING A NON-CASH EXPENSE OF $416,000 RELATED TO STOCK-BASED COMPENSATION.

Bioinformatics Notes

BIOMAX INFORMATICS will integrate a toxicogenomics database developed by the **MOUNT DESERT ISLAND BIOLOGICAL LABORATORY** with its data-management platform. The MDI database system includes information about cross-species interactions between genes, chemicals, and proteins that can be used to study disease susceptibility and diseases that are influenced by the environment.

SIMULATIONS PLUS has signed a multi-year collaboration with **ROCHE**, which will provide funding and feedback in developing the firm's GastroPlus software program. Simulations Plus will collaborate with Roche scientists to advance the capabilities of GastroPlus to simulate drug-drug interactions. Roche is slated to provide funding for the equivalent of one full-time scientist for two years.

FUNDED GRANTS

$1,200,000/FY 2008

EDAC: ENCODE DATA ANALYSIS CENTER

Grantee: Ewan Birney, European Bioinformatics Center

Began: May 15, 2008; Ends: Mar. 31, 2012

This proposal aims to facilitate the integration of data from multiple sources using sophisticated statistical models and machine learning techniques to build integration methods combining datasets. Birney and his team will also use this grant to provide quality assurance and summary metrics of genome-wide multiple alignments. Overall, they aim to provide deep integration of the ENCODE data, under the direction of the AWG and in tight collaboration with the other members of the ENCODE consortium.

$273,906/FY 2008

ADAPTIVE PERSONALIZED INFORMATION MANAGEMENT FOR BIOLOGISTS

Grantee: William Cohen, Carnegie Mellon University

Began: Jul. 11, 2008; Ends: May 31, 2012

This funding will enable the development of an adaptive information management tool. Cohen and his team intend to exploit recent advances in machine learning and database systems in order to facilitate their scheme for loosely integrating both structured information and unstructured text, and then querying the integrated information using easily formulated similarity queries.

Verifiable Protein Expression

37,000 tagged cDNA clones

TrueORF™

for tagged protein expression

TrueORF expresses the encoded transcript as an epitope-tagged protein, facilitating subsequent detection, purification and localization with anti-tag antibodies.

Genome-wide coverage: Human/mouse

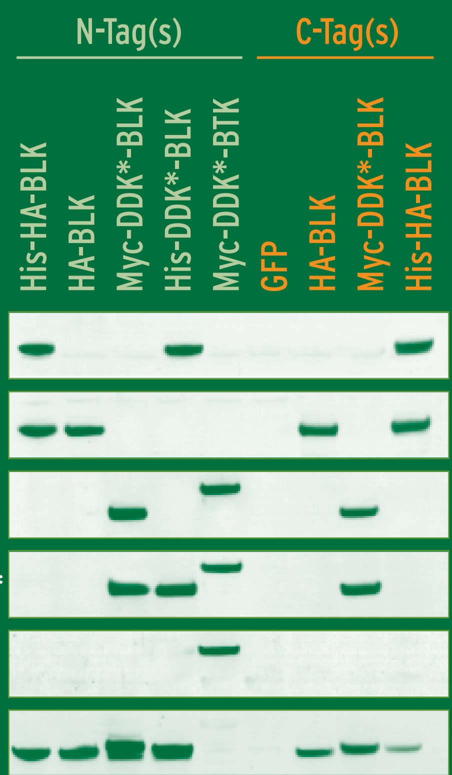
Sequence verified and guaranteed

The C-terminal dual tag of Myc and DDK*

Transfection-ready: Provided as 10 ug of purified plasmid

Easy shuttling into 20 tagged vectors using PrecisionShuttle™ system

5,000 clones ready for next-day delivery



The Western blot analysis of HEK293 cell lysate over-expressing BLK or BTK tagged with indicated epitopes.

*Peptide sequence of the DDK-tag (Flag®): N-DYKDDDDK-C
FLAG® is a registered trade mark of Sigma-Aldrich

origene.com/orf

BIOARRAYNEWS

For WTCCC, Agilent Will Provide Custom Chips

Agilent Technologies will provide custom, whole-genome copy number variation-focused arrays to the Wellcome Trust Case Control Consortium, which will use them in the second phase of its 19,000-sample study to identify genetic variants influencing disease susceptibility in a variety of rare and common diseases.

Specifically, the consortium will attempt to link genes to tuberculosis, coronary heart disease, types 1 and 2 diabetes, rheumatoid arthritis, Crohn's disease, bipolar disorder, autoimmune thyroid disease, ankylosing spondylitis, multiple sclerosis, breast cancer, and hypertension.

As part of its deal with the group, Agilent will design and fabricate the custom CNV chips, printing two arrays of 105,000 probes each per slide. Oxford Gene Technology, an Agilent certified service provider, will use Agilent's Velocity 11 Bravo robot to run the samples at OGT's lab.

For Agilent, the high-volume deal will enable it to hone a CNV-focused microarray product line scheduled to launch later this year. It also gives the Santa Clara, Calif.-based company the opportunity to iron out the protocol for high-throughput, automated array processing.

"We expect that enhancements to the comparative genomic hybridization workflow demonstrated in this project, such as the use of the Velocity 11 Bravo System and the use of a plate for purification, will facilitate widespread use of Agilent CGH/CNV microarrays in high-throughput environments," says Yvonne Linney, Agilent's vice president and general manager of genomics.

"Agilent also plans to provide a very high-resolution CNV-focused catalog array with 1 million features using data from public databases and Agilent's own and collaborative CNV research," she says.

Matthew Hurles, a geneticist at Wellcome Trust, said in a statement that the WTCCC "aims to characterize [the] most common structural modifications of DNA that may play a causative role in these diseases."

— *Justin Petrone*

DATAPOINT

$342

MILLION

ILLUMINA RAISED $342.6 MILLION IN PROCEEDS FROM ITS PUBLIC STOCK OFFERING THIS AUGUST.

Microarray Notes

DARPA has granted **COMBIMATRIX** $250,000 for proof-of-concept research to find new uses for the company's microarray technology, including label-free detection systems for use in diagnostics, chemical measurement, and chemical agent detection.

ASPER BIOTECH, an Estonian biotechnology company, has obtained a license to commercialize a new method for SNP-genotyping called arrayed primer extension-2 from the **ESTONIAN BIOCENTRE**.

Iowa-based **INTEGRATED DNA TECHNOLOGIES** has completed the expansion of its 21,500-square-foot European oligonucleotide production facility in Haasrode Research Park in Leuven, Belgium. The facility will allow it to provide better services for its customers in Europe, the Middle East, and Africa.

FUNDED GRANTS

$1,199,769/FY 2008

MICROARRAY CENTER FOR RESEARCH ON THE NERVOUS SYSTEM

Grantee: Dietrich Stephan, TGEN

Began: August 1, 2005; Ends: May 31, 2010

This Affymetrix Center of Excellence offers a platform to conduct research on the nervous system using microarrays. TGEN will provide assistance with experimental design, data generation, data interpretation, and data dissemination. New research goals include having scientists submit tutorials and projects online, review by an expert IRC, and data warehousing with a six-month timed release to the public.

$291,274/FY 2008

UNIVERSAL, COMPACT COMBINATORIAL MICROARRAYS FOR DNA BINDING SITE DISCOVERY

Grantee: Martha Bulyk, Brigham and Women's Hospital

Began: July 26, 2006; Ends: June 30, 2009

In this project, Bulyk will develop the use of compact combinatorial DNA microarrays in protein-binding microarray experiments in order to identify all possible DNA binding sites of sequence-specific transcription factors. She will also determine the binding affinities of all possible DNA binding sites for 15 *S. cerevisiae* transcription factors, as well as evaluate the utility of binding affinity data for improved prediction of *in vivo* binding sites.

Stand out from the background

It's no secret that background noise drowns out signals from low abundance proteins. SuperChip™ AB, the new Thermo Scientific microarray substrate for proteomics and discovery, features a proprietary agarose-based surface that significantly reduces the background that can interfere with signal detection.

Already familiar to scientists, agarose is an economical, protein-friendly substance that, when combined with our proprietary microarray surface, provides the covalent binding properties you demand for your research.

SuperChip AB is the clear choice to reduce background and provide a reliable surface for your protein studies.

Make sure all your results stand out. Choose SuperChip AB for spot-on results.

www.thermo.com/microarray
productinfo.microarray@thermofisher.com

SuperChip AB agarose-based microarray substrates for proteomic applications, enhance binding quality while reducing background.

Part of Thermo Fisher Scientific

Thermo
SCIENTIFIC

PHARMACOGENOMICSREPORTER

Pathwork First to Net FDA OK for CUP Dx

By being the first to launch an FDA- and CLIA-approved test designed to help physicians determine cancer type in a tumor, Pathwork Diagnostics may have gained a significant advantage over competitors.

This summer, Pathwork received 510(k) clearance from the FDA for its Tissue of Origin test.

The test, which uses Pathwork's PathChip microarray and runs on Affymetrix's GeneChip platform, "is the first custom Affymetrix gene expression array to be cleared for diagnostic use," the FDA noted in a statement.

Pathwork is not the only company with a diagnostic test for cancer of unknown primary origin. Currently, AviaraDx markets a CUP assay in the US, called Aviara CancerTYPE ID, which Agendia markets in Europe as CUP-Print. AviaraDx plans to submit its test for FDA clearance in the next 12 to 18 months.

Meanwhile, Rosetta Genomics and Exiqon are both developing a diagnostic for the indication.

However, Pathwork's product is the only test that is FDA-cleared for CUP, and its regulatory approval also marks an achievement for the agency. The Tissue of Origin test is only the second *in vitro* diagnostic multivariate index assay to receive the FDA's approval.

The FDA has said that it intends to regulate so-called IVDMIAs, a subset of laboratory-developed tests that combine multiple variables into a single, predictive test, and has issued two draft guidances in this regard. The first IVDMIA to receive a nod from the FDA was Agendia's MammaPrint assay for breast cancer recurrence.

According to the National Cancer Institute, between 2 percent and 4 percent of US cancer patients have a cancer for which the primary site is never identified. David Craford, Pathwork's VP of commercial operations, says that "the cost to identify a primary tumor can be significant," since traditional treatment approaches involve running multiple diagnostic technologies in parallel.

— *Turna Ray*

DATAPOINT

$27.8 MILLION

GENOMIC HEALTH REPORTED THAT SECOND-QUARTER REVENUE WAS $27.8 MILLION, UP 89.1 PERCENT FROM ITS REVENUE IN THE SAME QUARTER LAST YEAR.

PGx & Molecular Dx Notes

Northern Ireland-based **ALMAC DIAGNOSTICS** kicked off an international, multi-center collaborative lung cancer study aimed at finding a prognostic gene expression signature for patients with early non-small cell lung cancer. The study will be led by Dean Fennell at **QUEEN'S UNIVERSITY BELFAST**.

HX DIAGNOSTICS expanded a license and development agreement with **NANOGEN**, under which the companies will develop rapid, point-of-care diagnostics for respiratory infectious diseases. The firms are developing a fluid test to detect multiple strains and subtypes of influenza, including avian influenza, in one test.

SEQUENOM certified its first Center of Excellence for genotyping technology — the **MCGILL UNIVERSITY** and **GENOME QUEBEC** Innovation Centre.

FUNDED GRANTS

$658,595/FY 2008

PROTEOME SIGNATURES AND TARGET VALIDATION IN LYMPHOMAS

Grantee: Daniel Jay, Tufts University

Began: May 1, 2004; Ends: Jul. 31, 2010

Jay and his colleagues will use the funds to improve the use of proteomics as a pharmacogenomic tool by developing surface proteome signatures and performing functional proteomic target validation analysis directly on primary tumor tissue. Part of the proteome signature will help differentiate how someone is responding to a therapeutic.

$572,928/FY 2008

CORTICOSTEROID PHARMACOKINETICS AND PHARMACODYNAMICS

Grantee: William Jusko, SUNY Buffalo

Began: Jul. 1, 1977; Ends: Jun. 30, 2010

Currently, this NIGMS grant is funding Jusko and his colleagues to study pharmacogenomics as well as pharmacokinetics and pharmacodynamics in rat. The team will study genes from liver, muscle, and kidney tissue using microarrays, and hopes to assess the expression of biomarkers for diabetes.

Downstream | CASE STUDY

Diagnosing the Unknown

It's hard enough to treat cancer — but it's more difficult when the primary tumor can't be found. Researchers aim for molecular diagnostics to decipher the mystery of CUP. *By Ciara Curtin*

As a clinician, Gauri Varadharachary has seen quite a few patients with metastatic tumors for which no amount of tests or imaging could determine the primary tumor site. If after CT scans, PSA testing, mammograms, and more targeted testing, the site still remains elusive, then the patients are diagnosed with carcinoma of unknown primary and are often treated with systemic chemotherapy instead of targeted therapies. Varadharachary, though, is working to change that by developing molecular-based methods to determine the origin of these types of tumors.

These kinds of cancers are heterogeneous metastatic tumors. The incidence of CUP is estimated to be between three and five percent of all cancers, about the same percentage as pancreatic cancer. The American Cancer Society says that only half of CUP patients live nine to 12 months after their diagnosis. A 1994 study from the MD Anderson Cancer Center's James Abbruzzee, in whose group Varadharachary now works, reported 10 to 15 percent of CUP patients lived at least five years after diagnosis.

Recently, Varadharachary and her colleagues in the Abbruzzee group have started focusing on determining molecular profiles of subsets of CUP so that a tailored treatment can be developed to better treat them. "In the era of molecular diagnostics, it's the right time to define and select subtypes," she says. In particular, the team is using immunohistochemistry and RT-PCR panels.

In a recent *Lancet Oncology* paper, Varadharachary reports that she's been able to parse out a subtype of CUP based upon its immunohistochemistry. By looking at the cytokeratins on the tumor cell surface — which aren't altered as a cell transforms from normal to malignant — Varadharachary can trace the cell's lineage back to before it became a tumor. When a tumor has a CK20+/CK7- immunostain profile and is positive for CDX2, a nuclear transcription factor and product of a homeobox gene that promotes intestinal differentiation, Varadharachary calls this colon-cancer-profile CUP. The patients with this profile, she says, benefit from colon-cancer-based treatments — a sign that it really was a metastatic colon tumor.

In addition to immunohistochemistry, Varadharachary is also working on a molecular assay to diagnose CUP subtypes. She and her colleagues are developing a prospective trial of a 10-gene RT-PCR panel. They're including colon-specific marker cadherin-17, which is also expressed in both normal and cancerous colon tissues. They will then be testing the panel on formalin-fixed paraffin-embedded tissue samples from CUP patients — a technical feat that spares patients additional biopsies for fresh sample.

Validating that those subtypes are what they suspect, though, is a challenge. By definition, no one knows from what tissue the CUP tumor came. "Validation is indirect," Varadharachary says. She adds that it is based on how long the patients live, how aggressive their cancer is, and how they react to a tailored treatment, if one exists. If the tumor has the markings of a colon cancer and also acts like it, then it probably is colon cancer.

But will molecular diagnostics be able to determine CUP tissue of origin? "It's still early to say," Varadharachary says. The tests will work, she says, but not alone. "There won't be one gold standard," she says; molecular diagnostics will have to be used in conjunction with imaging and immunohistochemical assays. Part of the ongoing project is to assess how the RT-PCR molecular diagnostic correlates with the immunohistochemistry results.

In the end, determining the CUP's original tissue comes down to patient care. By knowing someone has colon cancer rather than breast cancer, there are different therapies to give the patient a better chance of survival. But there's a dearth of tailored treatments for many cancers, such as pancreatic cancer, Varadharachary says. When that changes, treating CUP patients will improve as well.

Too Much Information

Clinical genomics advisor Gurvaneet Randhawa on evidence-based medicine and the 'information overload' that's making it tough to translate research into patient care.

Gurvaneet Randhawa is the senior advisor on clinical genomics and personalized medicine at the Agency for Healthcare Research and Quality, a unit within the US Department of Health and Human Services. GT's Jeanene Swanson caught up with him to discuss the use of evidence-based medicine in the clinic, and how far along systems biology tools have come in affecting patient care.

GENOME TECHNOLOGY: What is evidence-based medicine?

GURVANEET RANDHAWA: Evidence-based medicine typically implies three different elements that need to be considered together, the first being the best and current scientific evidence, the second being the clinical expertise, and the third being the patients' values — combining all of these in making decisions jointly between the clinician and the patient in terms of what is the best course of action to take.

GT: How do you apply that in clinical practice?

GR: That is a very challenging aspect, primarily because given the workload, there really isn't enough time for the busy primary care physician to look up all the latest studies that are being published, let alone try and synthesize them in a fashion that is digestible and useable in the clinic. To do all of those steps — specifying what questions you want to answer, collecting all of the research evidence in a systematic fashion, and integrating that knowledge with expertise and patient values — requires a multidisciplinary team of people working on these reports for several months.

The challenge we are struggling with is how to do it more consistently and uniformly, and one of the options is to have available different credible sources of information that can guide the clinicians. An example of that are the recent recommendations released by the US Preventive Services Task Force on prostate cancer screening.

GT: How does large-scale biological research play a role here?

GR: The task force hasn't ventured too much in that area because it deals mostly with clinical prevention.

AHRQ's mission is to improve the effectiveness, safety, quality, and efficiency of healthcare. So our focus is on things that are already being used in clinical practice, or that are new to clinical practice and still haven't gained widespread use. All of the 'omics studies tend to be more exploratory analyses.

GT: What are some of the challenges facing evidence-based medicine?

GURVANEET RANDHAWA

GR: Getting good, validated information is the first challenge. It's important to have large-scale studies, but ... there are many other factors beyond genetics and beyond markers that influence the causation of disease.

The second challenge is, even if we know what genes or what biological factors are predictive of what diseases, what do we do with that information? There is no drug that is without adverse events. We have to assess the balance of benefits and harms.

The third challenge is trying to implement the best possible research evidence into practice, and again, there's a huge information overload that is occurring. The information needs to be reliable and credible, and you have to make sure that it is interoperable across different systems.

Events | MEETINGS AND DEADLINES

Conferences

DATE	CONFERENCE	ORGANIZER	LOCATION	CATEGORY
Sep 1-3	Protein Function Prediction Tools Workshop	EBI-ENFIN	Hinxton, UK	Bioinformatics
Sep 1-4	11th International Meeting of MGED	Microarray and Gene Expression Data Society	Riva del Garda, Italy	Arrays
Sep 2-6	Metabolomics 2008	Metabolomics Society	Boston	Metabolomics
Sep 7-10	AIRI Annual Meeting	Association of Independent Research Institutes	Washington, DC	General
Sep 10-14	Genome Informatics	CSHL/Wellcome Trust	Hinxton, UK	Bioinformatics
Sep 16-18	RNAi Europe	Select Biosciences	Stockholm, Sweden	RNAi
Sep 17-18	Advances in qPCR	Select Biosciences	Stockholm	PCR
Sep 22-23	Biosimilars2008	Scherago International	Washington, DC	Clinical
Sep 24-25	Exploring Next-Generation Sequencing	CHI	Providence, RI	Sequencing
Sep 25-28	The Pathologists' Meeting	College of American Pathologists	San Diego	Clinical
Sep 27-30	HGM 2008: Human Genome Meeting	HUGO	Hyderabad, India	Genomics
Sep 29 - Oct 1	Biomarker Discovery Summit	CHI	Philadelphia	Biomarkers
Oct 9-12	Personal Genomes	Cold Spring Harbor Laboratory	Cold Spring Harbor, NY	Translational
Oct 12-17	International Biotechnology Symposium & Exhibition	IUPAC	Dalian, China	General
Oct 14-17	NIH Research Festival	NIH	Bethesda, Md.	General
Oct 16-17	Personalized Health Care National Conference	Ohio State University	Columbus, Ohio	Translational
Oct 20-22	Discovery2Diagnostics	IBC	San Diego	Translational
Oct 20-24	Discovery on Target	CHI	Boston	Pharma
Oct 22-24	Northeast Regional Life Science Core Directors Meeting		Burlington, VT	Core labs/ instrumentation
Oct 30 - Nov 2	AMP 2008	Assocation of Molecular Pathologists	Grapevine, Texas	Clinical
Nov 1-5	Mass Spectrometry Applications to the Clinical Laboratory	University of California, San Diego	San Diego	Proteomics
Nov 10-11	Burrill Personalized Medicine Meeting	Burrill & Company	San Francisco	Personalized medicine
Nov 10-13	qPCR Symposium USA	IES / TATAA	Millbrae, Calif.	PCR
Nov 11-15	ASHG 2008	American Society of Human Genetics	Philadelphia	Genomics
Nov 13-14	Personalized Medicine Conference	Harvard Partners Center	Boston	Personalized medicine
Nov 15-19	Neuroscience 2008 Annual Meeting	Society for Neuroscience	Washington, DC	Neuroscience
Dec 13-17	American Society for Cell Biology Annual Meeting	ASCB	San Francisco	General
2009				
Jan 5-9	Pacific Symposium on Biocomputing		Kohala Coast, Hawaii	Bioinformatics
Jan 10-14	Plant and Animal Genome Meeting XVII	Scherago	San Diego	Genomics
Jan 11-16	PepTalk	CHI	San Diego	Proteomics
Feb 4-7	Advances in Genome Biology and Technology	Gcorp	Marco Island, Fla.	Genomics
Feb 7-10	ABRF 2009	Association of Biomolecular Resource Facilities	Memphis, Tenn.	Core facilities
Feb 12-14	AUTM 2009	Association of University Technology Managers	Orlando	Tech transfer
Feb 22-25	Fifth Annual US HUPO	US HUPO	San Diego	Proteomics
Feb 24-27	Molecular Medicine Tri Conference	CHI	San Francisco	Translational

Deadlines

SEPTEMBER 8
Application deadline for NIH's **SHARED NEUROBIOLOGY OF FRAGILE X SYNDROME AND AUTISM** grant, which will award funds to study the neurobiology of patients with both FXS and autism and to identify novel drug targets for these diseases.

SEPTEMBER 8
Application deadline for the **DIRECTED STEM CELL DIFFERENTIATION FOR CELL-BASED THERAPIES FOR HEART, LUNG, AND BLOOD, AND AGING DISEASES** grant. This NIH award is for researchers to study the differentiation of embryonic or adult stem or progenitor cells, either *in vitro* or *in vivo*.

SEPTEMBER 8
Application deadline for the **PILOT STUDIES IN PANCREATIC CANCER** grant. This funding from NIH will support multi-disciplinary research of pancreatic cancer, including studies on genetic causes, biomarkers, and the identification of new drug targets.

SEPTEMBER 8
Application deadline for the **DEVELOPMENT OF BIOMARKERS FOR MENTAL HEALTH RESEARCH AND CLINICAL UTILITIES** grant from NIH. This award supports research and development of commercializable biomarker technologies relevant to mental disorders.

SEPTEMBER 22
Application deadline for the **CISE COMPUTING RESEARCH INFRASTRUCTURE** program. This NSF award will fund the creation, enhancement, and operation of world-class computing research infrastructure.

SEPTEMBER 25
Application deadline for the NIH award for **APPLICATION OF EMERGING TECHNOLOGIES FOR CANCER RESEARCH**. This grant will fund the development of technologies, including tools and instrumentation, for cancer-relevant molecular analyses *in vitro*, *in situ*, and/or *in vivo*. This grant is part of the broader NCI-sponsored Innovative Molecular Analysis Technologies program.

SEPTEMBER 25
Application deadline for the **INNOVATIVE TECHNOLOGY SOLUTIONS TO CANCER SAMPLE PREPARATION** grant. This NIH award, also part of the NCI's IMAT program, will support technology development in preparing, purifying, processing, and handling cancer-relevant samples for molecular analyses.

Qualify First

MassARRAY® System for Copy Number Analysis – The strongest in fine mapping.

- Validate whole genome array data
- Increase sample throughput
- Determine absolute copy number
- Discover allele-specific inheritance patterns

What do the Tour de France and MassARRAY CNV analysis have in common? Find the answer at www.sequenom.com/cycling

SEQUENOM®
www.sequenom.com

sq110 - CNV - GEN Half - 08/08

Lab Reunion | THE ALTMAN LAB

A Man of Many Hats

Russ Altman is many things to many people: researcher, physician, academic advisor. He's never lost sight of his passion for science and the high expectation he places on himself and those who pass through his lab. *By Matthew Dublin*

As an MD/PhD, Russ Altman is a member of that rare breed of individuals who really do have one foot planted firmly at the bench and the other in the clinic. Upon finishing his undergraduate studies in biochemistry at Harvard University in 1983, he headed off to Stanford University, initially thinking he would pursue graduate work in biophysics. But when he arrived, he heard about the university's medical information sciences program that appealed to his twin interests of computer science and biology. Years later, when Altman took over as director of the program in 2000, he renamed it the biomedical informatics program. Also the chair of Stanford's bioengineering department, Altman continues to maintain a small medical practice despite juggling his many research projects.

Altman's lab is currently focused on developing PharmGKB, a pharmacogenomics database that aims to be the Web destination for information about genes that are important for drug response. "We're building this database and we're doing informatics research on how to deliver services and how to analyze data relevant to pharmacogenomics," he says. He also directs Stanford's Helix Group, which uses simulation, machine learning, and natural language processing methods to conduct protein and RNA structure analysis. And as principal investigator for the NIH Center for Biomedical Computation at Stanford, Altman helps develop software tools to simulate biological systems in terms of the motions of their components.

In addition to dual roles of researcher and physician, Altman is also a mentor and advisor to many. He says his approach to advising is based upon what he considers to be an "old fashioned" interpretation of the PhD as a hunting license to define and solve a problem independently. Along those lines, he tries to stay as much out of his students' way as possible during the early stages of their journey to a doctorate. "I try not be too directive in the very beginning stages where they're looking for their thesis project. I have a ton of ideas but I really feel it's important for them to articulate what the problem is, why it's important, and how to go after it," Altman says. "I work with them on all of those things, but it would be too easy for me to feed them the problem and why it's important — and then they would just be technicians implementing the work. … That might be fine for getting a PhD, but then when they hit the real world, they have not done this activity that the world is expecting them to do."

> **"I try not to be too directive in the very beginning stages where they're looking for their thesis project."**

What they're made of

Altman says that his grad students really get to see what they're made of in year two or three of their graduate studies. "They have done the confidence builder projects, maybe they've been on one paper as a middle author, but now they have to commit themselves to three or four straight years working very deeply on a problem and defining what it is," he says. "It's hard, and I want them to be primarily responsible for it. And they are very stressed out, more than in any other time." Sometimes, he must act as a life preserver for a student by throwing in his own ideas about the direction their research should take, but this rarely yields the desired outcome of producing an independent and mature investigator.

Pressure and dedication take on a whole new meaning for those brave souls endowed with the

desire to embark on the MD/PhD path to which Altman was drawn. He has some advice for those considering it: "The first thing I say to somebody like that is, 'You have to want to be a doctor to go to medical school, and you have to want to be scientist to go to PhD school,'" he says. "What I mean by that is: do not do one simply to give you a competitive advantage in the other, because the training for both are long and hard, and any kind of superficial reason to do it cannot weather how physically and intellectually difficult it is to finish them." And when it comes to looking for the right MD/PhD program, Altman says that the quality of the PhD training should be the deciding factor. "Medical school is basically the same for everybody. At the end of the day, everyone takes the same nationalized exam, so the quality of medical school training is pretty much even," says Altman. "But the quality of PhD training is highly variable. Given that, I would choose your program based on offering first-class PhD credentials and then the MD program."

Former Altman postdoc Alain Laederach, now a research scientist at the Wadsworth Center for Developmental Genetics and Bioinformatics, says his favorite memory is a rather terrifying experience involving the grant proposal for the Simbios group within Altman's NIH center, which was around 250 pages long. "Three days before it was due, Russ decided we needed to completely reorganize the structure of the grant," says Laederach. "It was exactly what the grant needed, but it took a lot of guts to start cutting and pasting everything in such a major way three days before the due date. But we ended up getting it funded."

Laederach says that he learned from Altman's approach to getting the most out of the latest computational approaches to solve real-world biological problems. "In Russ's lab, I learned to identify biological problems that are tractable computationally and identify the best approach," says Laederach. "I also learned to identify the medical applications of the biology, which is critical to obtaining funding from NIH."

>NAMING NAMES

Any number of scientific greats have been shaped over the years in Altman's lab. Here are just a few of those who earned their stripes with him.

JEFF CHANG
According to Altman, Chang arrived in his lab with coding skills at a level he has yet to see since. Chang actually joined the lab as a sophomore and later proceeded to write a "trail blazing" thesis on natural language text mining, Altman says. He is currently gearing up for a postdoc at the Duke Institute for Genome Sciences and Policy.

RICH CHEN
Chen joined the lab as an MD student originally intending to do a short rotation as a researcher. But, as fate would have it, a local Silicon Valley venture capitalist happened to see a presentation he and fellow labmate Ramon Felciano gave and offered the students several million dollars to start up the company now known as Ingenuity.

RAMON FELCIANO
During his stint with Altman, Felciano co-founded Ingenuity, where he acts as both chief technology officer and vice president of R&D. He has also led the development of many informatics and Web-based projects, including RiboWeb, a semantic application for Internet-based, collaborative molecular biology.

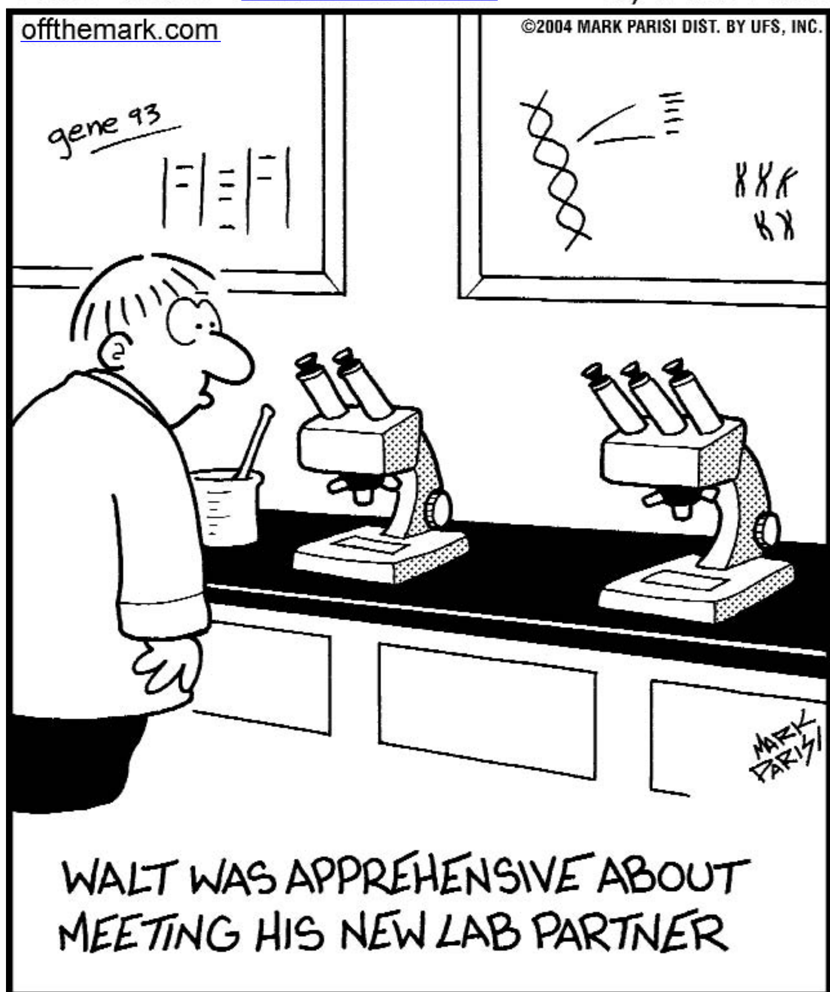
SEAN MOONEY
Mooney arrived in Altman's lab as a postdoc and started BioE2E, an organization geared toward grad students and postdocs interested in the entrepreneurial side of biotechnology. He is currently an assistant professor in the Department of Medical and Molecular Genetics at the Indiana University School of Medicine.

SOUMYA RAYCHAUDHURI
During his stint at Altman's lab, this former MD/PhD student published nine papers, the record for the most papers by any student in the lab. Currently, Raychaudhuri is a research fellow in the program of medical and population genetics at the Broad Institute.

OLGA TROYANSKAYA
Troyanskaya says Altman taught her to be suspicious of her results and how to pick the best collaborators, among many other lessons. Currently, she is an assistant professor at Princeton University's Lewis-Sigler Institute for Integrative Genomics, where she works on computational techniques for genomic data integration, microarray analysis, and pathway identificiation.

Blunt End

HUMOR, WE HOPE



BPA WORLDWIDE BUSINESS
Genome Technology (ISSN 1530-7107) is published ten times a year (monthly except combined issues in Jul/Aug and Dec/Jan) by GenomeWeb, LLC, 125 Maiden Lane, New York, NY 10038. Periodicals postage paid at New York, NY, and additional mailing offices. Genome Technology is sent free of charge to qualified professionals in life sciences research. Non-qualified rate is $149 per year. POSTMASTER: Send address changes to Genome Technology, Peck Slip Station, PO Box 998, New York, NY 10273.

Discovery on TARGET

October 20-23, 2008
World Trade Center • Boston, MA
Exhibits: October 21-23, 2008

The only dedicated "Hot Targets" event!

October 21-22

RNAi for Screening
How to Best Utilize RNAi as a Screening Tool

Kinase Inhibitors
Moving Forward into Clinical Studies

Ion Channels as Therapeutic Targets
A Flood of Potential for Drug Discovery

October 22-23

RNAi for Therapeutics
How to Best Transition From the Lab to the Clinic

HDAC Inhibitors
Targeting Oncology and Beyond

Targeting Diabetes with Novel Therapeutics
New Drug Targets Expand Treatment Options

Short Courses:

- (SC1) Imaging Technologies for Target Discovery
- (SC2) Understanding the Structural Biology of Ion Channels to Guide Drug Discovery
- (SC3) Advances in DNA Methylation Analysis
- (SC4) Tackling RNAi Delivery
- (SC5) Ion Channel Assays for Safety Screening
- (SC6) Screening For Potential Drug Targets - Design Strategies for Novel-Generation Kinase Inhibitors

Highlighted Speakers:

Thomas Singer D.A.B.T. Global Head of Non-Clinical Safety, F. Hoffmann-La Roche AG
Elena Feinstein, M.D., Chief Scientific Officer, Quark Pharmaceuticals Inc.
Steven C. Quay, M.D., Ph.D., Chief Scientific Officer, MDRNA Inc.
Tod Woolf, Ph.D., President and Chief Executive Officer, RXi Pharmaceuticals
Klaus Giese, Ph.D., Chief Scientific Officer, Silence Therapeutics PLC
George K. Chandy, M.D., Ph.D., Professor, Physiology & BioPhysics, School of Medicine, University of California - Irvine
Michael Dabrowski, Ph.D., Head of Global Ion Channel Initiative, AstraZeneca
Laszlo Kiss, Ph.D., Senior Research Fellow, Merck Research Laboratories
William C. Shakespeare, Ph.D., Senior Director, Chemistry, Ariad
Jurgen Wagner, Ph.D., Director, G.D.C., Novartis, Institutes for Biomedical Research

Organized By:

Cambridge Healthtech Institute
250 First Avenue, Suite 300
Needham, Massachusetts 02494
Telephone: 781-972-5400 or
Toll-Free in the U.S. 888-999-6288
Fax: 781-972-5425 • www.healthtech.com

Make sure to mention key code o30 when registering!

www.DiscoveryOnTarget.com

1000-Series Thermal Cyclers

YIELD

SPEED

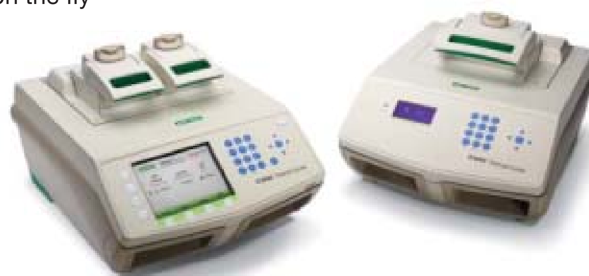
WE THINK YOU SHOULD BE ABLE TO OPTIMIZE ON THE FLY

Convenience Is What Happens When You Rethink PCR

Change the way you think about PCR with Bio-Rad's new family of thermal cyclers.

Wouldn't you rather optimize your reactions in minutes and not days? With Bio-Rad's new 1000-series thermal cyclers, optimizing on the fly is just the beginning.

- Easily interchangeable reaction modules meet any experimental or throughput need
- Reduced-mass sample blocks increase ramp rates and reduce run times
- Thermal gradient lets you incubate each row at a different temperature for fast protocol optimization



When you rethink PCR, you realize how easy it should be.

For more information, visit us on the Web at **www.bio-rad.com/pcr**

Purchase of this instrument conveys a limited non-transferable immunity from suit for the purchaser's own internal research and development and for use in applied fields other than Human In Vitro Diagnostics under one or more of U.S. Patents Nos. 5,656,493, 5,333,675, 5,475,610 (claims 1, 44, 158, 160–163 and 167 only), and 6,703,236 (claims 1–7 only), or corresponding claims in their non-U.S. counterparts, owned by Applera Corporation. No right is conveyed expressly, by implication or by estoppel under any other patent claim, such as claims to apparatus, reagents, kits, or methods such as 5' nuclease methods. Further information on purchasing licenses may be obtained by contacting the Director of Licensing, Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404, USA.

To find your local sales office, visit **www.bio-rad.com/contact/**
In the U.S., call toll free at 1-800-4BIORAD (1-800-424-6723)

Visit us on the Web at discover.bio-rad.com

BIO-RAD