CAMBRIDGE HEALTHTECH MEDIA GROUP
Bio•IT World
Indispensable Technologies Driving Discovery, Development, and Clinical Trials

Welcome to your digital edition of *Bio•IT World* magazine

We hope the publication is meeting your expectations! ***Bio•IT World*** continues to build stronger relationships with readers by understanding and addressing career and business needs in editorial content.

As a subscriber to ***Bio•IT World,*** the flagship publication of Cambridge Healthtech Institute(CHI), you also have complete access to an integrated *Life Science Network,* including:

- Lead generation programs
- Insight Pharma Reports
- CHI Conferences
- Professional services (Cambridge Marketing Consultants, Cambridge Meeting Planners, Public Service Announcements)
- Cambridge Healthtech Media Group offering five topic-specific eNewsletters, whitepapers, webcasts, podcasts, and web sites — bio-itworld.com, pharmaweek.com, digitalhcp.com — dedicated to keeping you informed on all aspects of the life sciences industry ***and***
- Barnett Educational Services, offering seminars, training strategy and consulting, and industry-leading reference manuals and publications.

CHI is the preeminent *Life Science Network* for leading researchers and business experts from top pharmaceutical, biotech and academic organizations. We strive to develop quality information resources that provide valuable new insights and competing points of view while offering balanced coverage of the latest life science developments.

As always, we welcome any and all comments or suggestions.

>>Click Here for technical assistance with your digital magazine.
>>Click Here for all other subscription customer service requests.

Bio-IT World reports the following digital circulation figures as verified by the Qmags Self-Audit procedures:

October 2008 Issue	
Total qualified digital subscribers:	7,689
Biopharma/Chemicals	2,080
Biotech/Bioinformatics	2,081
Acad/Gov, Clinical Trial, CRO	2,090
Hosp Lab/Healthcare, Other	1,438
September 2008 Issue	
Total emails sent:	7,516
Undeliverable :	67
Total delivered:	7,443
Removed:	34

NAVIGATION KEY

- Previous/Next Page: mouse click (or keyboard) on left/right arrows to turn page
- Contents: takes you to theTable of Contents page
- Zoom In: top will zoom in to the top of page & bottom will zoom in to the bottom of page
- Zoom Out: takes you from single to double page view
- Front Cover: takes you to the Cover Page of magazine
- Search Issue: allows full search by keyword or phrase (can also search archives!)

I WANT

TO WORK ANYWHERE I CHOOSE

First backlit keyboard for business.
Dell Connection Manager. Up to 19 hours of battery life.

THE NEW LATITUDE™ E6400

DELL.COM/IWantSimplicity 800.727.4247

Cambridge Healthtech Media Group

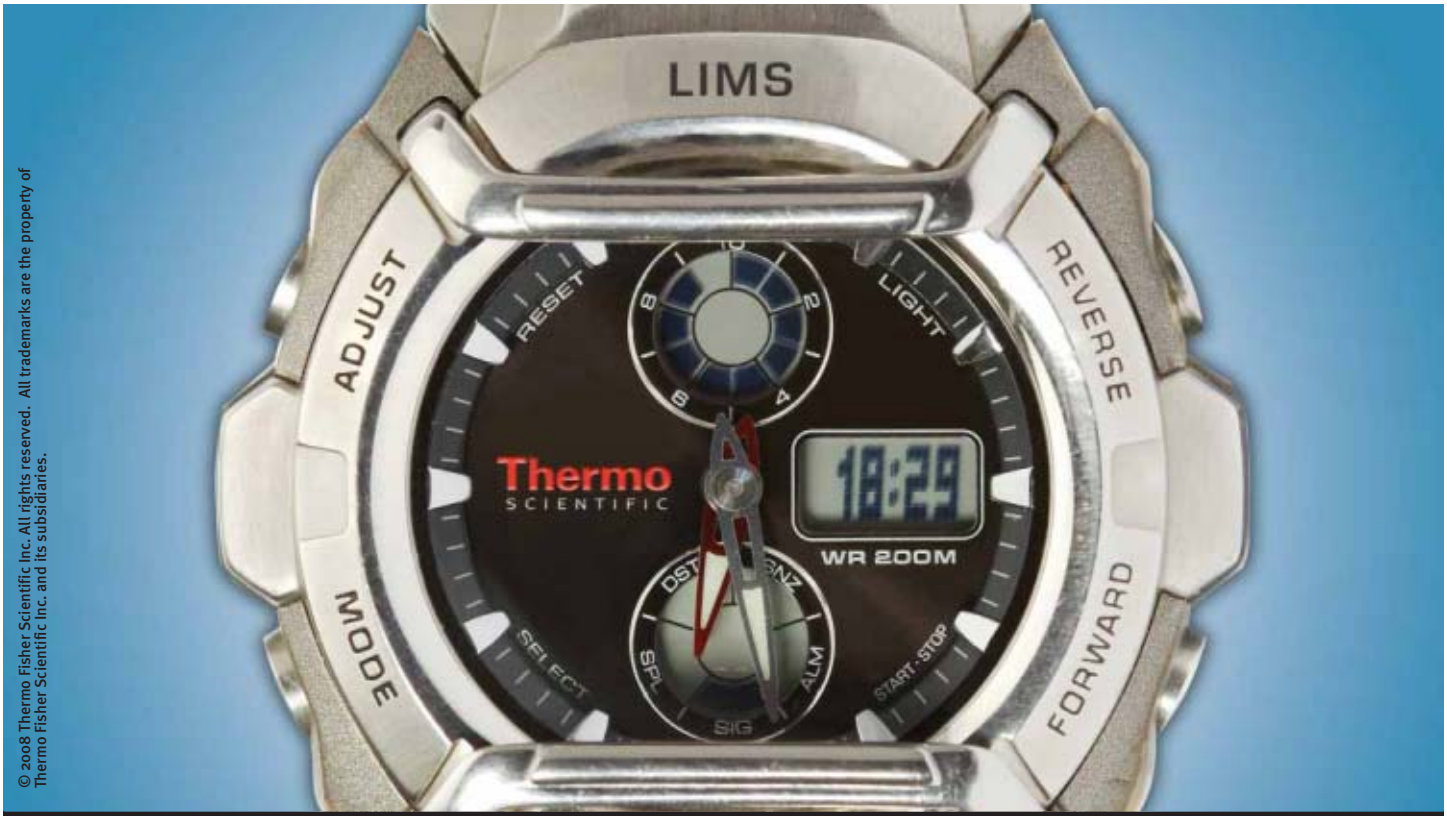
www.bio-itworld.com

Bio·IT World

Indispensable Technologies Driving Discovery, Development, and Clinical Trials

OCTOBER 2008 • VOL. 7, NO. 8

Eric Schadt's Integrative Approach to Predictive Biology 20



It's about compliance.

Which is why it's about time you invested in a Thermo Scientific LIMS. Compliance requirements are built into the core of every Thermo Scientific LIMS, giving you confidence in your regulatory submissions, improving your delivery time and margins. The more you reduce your exposure to risk in the delivery chain, the more control you'll have over your end product. Which means greater consumer confidence in your product and shareholder confidence in your business. No wonder more laboratories count on Thermo Scientific LIMS than any other.

Now with Thermo Scientific WebAccess Suite,™ all you need is a web browser for full, secure, high performance access to any Thermo Scientific informatics solution – from anywhere. Web deployment of any Thermo Scientific LIMS. That's all about compliance. Isn't it about time? Your time.

+1 866-463-6522 • marketing.informatics@thermofisher.com

FREE DEMO. Thermo Scientific WebAccess Suite™ - the ultimate source of integration for your lab and for your business. Get instant transparent integration of people, instruments and data, giving you secure and reliable efficiencies and compliance across your enterprise.

For more information and a FREE DEMO, visit www.thermo.com/webaccess

Moving science forward

Thermo SCIENTIFIC

Part of Thermo Fisher Scientific

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:

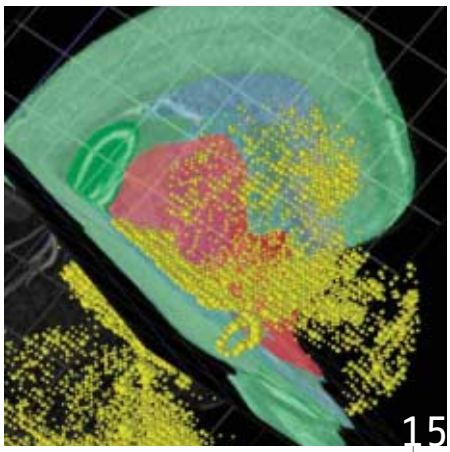
- Simplified primer design and virtual cloning, thanks to the added primer designing capabilities in SeqBuilder
- Faster and easier SNP detection and analysis for your specific project needs
- Ideal for Next Generation platforms; can be used with SeqMan NGen for enhanced analysis and visualization of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

DNASTAR® Lasergene

Contents [10·08]



Cover Story

Eric Schadt's Integrative Approach to Predictive Biology

The difference is in the layers. Eric Schadt believes that genome-wide association studies are just the beginning. 20

First Base

Circuit City: Genes and Diseases

KEVIN DAVIES

There was much activity at the Broad Institute last month. First there was a ceremony to mark the historic pledge of $400 million from billionaire philanthropists Eli and Edythe Broad to provide an endowment for the institute that guarantees its future (not that there was really any doubt—see page 8).

Two days later, 300 scientists gathered for the second Wellcome Trust/*Nature Genetics* symposium on the Genomics of Common Diseases. I'm not going to detail any of the highlights here—the meeting followed Gordon Conference rules designed to encourage open presentation and discussion of unpublished results. Suffice it to say that the rapid progress in defining the genes and copy number variants (CNVs) responsible for common diseases such as cancer, heart disease, diabetes, and mental illness continues apace.

The biggest splash in genome-wide association studies (GWAS) in 2007 was made by the Wellcome Trust Case Control Consortium (WTCCC), which identified dozens of gene variants for seven common diseases. In some cases, such as Crohn's disease, more than 30 confirmed gene associations have now been reported. This summer, a WTCCC sequel was announced, in which consortia will use new technology from Illumina, Affymetrix, and Agilent to analyze 120,000 DNA samples. This project, considered the world's largest genetic research study in history, will include both SNPs and CNVs in 27 diseases such as multiple sclerosis, asthma, and schizophrenia.

But simply compiling lengthy lists of genes using GWAS and resequencing targeted regions (see page 13) won't be enough. For a start, GWAS only provides the power to detect relatively common gene variants. A growing consensus is that rare variants may cumulatively explain much of the unexplained genetic variance in common disease susceptibility. Indeed, several speakers at the Broad meeting presented examples of sequencing projects in which mutations of rare or intermediate frequency have been uncovered. The deCODE group in Iceland recently published a method that it believes will help discover rare disease-causing variants, at least until we get closer to the $1000 genome (Kong, A. *et al. Nature Gen.* Sept 2008).

The other challenge is how to translate lists of seemingly unrelated gene-disease associations into useful diagnostics or therapeutics. Few groups have made as much of an impact in this arena lately as Eric Schadt's team at Merck (he's part of the Rosetta Inpharmatics subsidiary in Seattle)—the subject of this month's cover story (see page 20). Schadt's team is leading a multidimensional approach to tackling common diseases.

Rather than just look at genetics or gene expression, Schadt integrates these data streams and others. As detailed in our story, in at least one published case, his team showed that one of the WTCCC disease-associated genes was likely to be incorrect. More importantly, Schadt considers the assessment and modulation of specific gene circuits rather than individual genes. (You can read John Russell's entire interview with Schadt in the inaugural issue of *Predictive Biomedicine*, the latest *Bio•IT World* eNewsletter)

Several complementary approaches are also being tried to make sense of the GWAS bounty, including functional genomics methods that look at protein-protein interactions, as well as text mining of the scientific literature. Various methods of capturing coding genes and resequencing also offers exciting opportunities, as does improved understanding of the extent of CNVs.

Such methods will prove invaluable in unraveling the genetics of cancer. The latest tour de force analysis from Johns Hopkins' Bert Vogelstein, Victor Velculescu, and Ken Kinzler (see page 34) have documented hundreds of mutated genes in dozens of pancreatic and brain cancer samples (echoing findings in breast and colon cancer reported in 2007). The average tumor cell carries more than 60 genetic alterations, with one gene mutation seemingly distinguishing different severities of brain cancer. Nevertheless, the authors conclude that the best route to new therapies probably lies in targeting "the physiologic effects of the altered pathways and processes, rather than their individual genetic components."

Company Index

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

PS Form 3526-R

1. **Title of publication:** Bio-IT World
2. **Publication Number:** 1538-5728
3. **Filing Date:** 10/1/08
4. **Issue Frequency:** monthly except for July/Aug and Dec/Jan
5. **Number of Issues Published Annually:** 10
6. **Annual Subscription Price:** $199 US
7. **Mailing Address of Publication:** Bio-IT World, 250 First Ave, Suite 300, Needham, MA 02494
   Contact Person: JoAnne M Sperry (781) 972-1350
8. **Mailing Address of Headquarters:** (same address)
9. **President:** Phillips Kuhl, **Editor-In-Chief:** Kevin Davies, **Executive Editor:** John Russell (same address)
10. **Owner:** Cambridge Bio Collaborative, 250 First Ave, Suite 300, Needham, MA 02494, Phillips Kuhl

		Average No. Copies Each Issue During Preceding 12 months	No. Copies of Single Issue Published Nearest to Filing Date
Total Number of Copies			
15a.	Total Press Run	23,306	21,559
15b(1).	Paid/Requested Distribution by USPS:	20,417	20,055
15b(3).	Paid/Requested Distribution Outside USPS:	78	37
15c.	Total Paid/Requested Circulation*:	20,495	20,092
15d(1).	Nonrequest Distribution by USPS:	1,929	1,044
15d(4).	Nonrequest Distribution Outside USPS:	619	100
15e.	Total Nonrequest Distribution	2,548	1,144
15f.	Total Distribution:	23,043	21,236
15g.	Copies Not Distributed:	286	323
15h.	Total:	23,329	21,559
15i.	Percent Paid/Requested:	88.9%	94.6%

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on this form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

JoAnne M Sperry, Circulation & Marketing Manager, Bio-IT World

*All Bio-ITWorld distribution is outside County. Due to US Postal regulations, Paid/Qualified requests for digital editions are not included in the numbers above. Total digital distribution averages 6,396 additional copies each issue.

VOLUME 7, NO. 8

Editorial, Advertising, and Business Offices: 250 First Avenue, Suite 300, Needham, MA 02494; (781) 972-5400

Bio IT World (ISSN 1538-5728) is published monthly except July/August and December/January by Cambridge Bio Collaborative, 250 First Avenue, Suite 300, Needham, MA 02494. *Bio IT World* is free to qualified life science professionals. Periodicals postage paid at Boston, MA, and at additional post offices. The one-year subscription rate is $199 in the U.S., $240 in Canada, and $320 in all other countries (payable in U.S. funds on a U.S. bank only).

POSTMASTER: Send change of address to Bio-IT World, P.O. Box 3414, Northbrook, IL 60065. Canadian Publications Agreement Number 41318023. CANADIAN POSTMASTER: Please return undeliverables to Station A, P.O. Box 12, Windsor, ON N9A 6J5.

Subscriptions: Address inquires to Bio-IT World, P.O. Box 3414, Northbrook, IL 60065 (888) 835-7302 or e-mail biw@omeda.com.

CAMBRIDGE HEALTHTECH MEDIA GROUP

Bio·IT World®

Indispensable Technologies Driving Discovery, Development, and Clinical Trials

EDITOR-IN-CHIEF
Kevin Davies (781) 972-1341
kevin_davies@bio-itworld.com

EXECUTIVE EDITOR
John Russell (781) 972-1342
john_russell@bio-itworld.com

MANAGING EDITOR
Allison Proffitt (781) 972-1345
aproffitt@healthtech.com

ART DIRECTOR
Mark Gabrenya (781) 972-1349
mark_gabrenya@bio-itworld.com

WEB EDITOR
Catherine Varmazis (781) 972-1344
catherine_varmazis@bio-itworld.com

DIRECTOR, BUSINESS DEVELOPMENT
Angela Parsons (781) 972-5467
aparsons@healthtech.com

VP SALES – CA, WESTERN US, CANADA, PACIFIC RIM
Alan El Faye (213) 300-3886
alan_elfaye@bio-itworld.com

REGIONAL SALES MANAGER – NEW ENGLAND, NORTH EASTERN US
Kay O. Christopher (860) 693-2991
kchristopher@healthtech.com

REGIONAL SALES MANAGER – MIDWEST, SOUTH EASTERN US, EUROPE
Tom Loughran (781) 972-5482
tloughran@healthtech.com

SENIOR DIRECTOR OF MARKETING & OPERATIONS, PUBLICATIONS
Joan A. Chambers (781) 972-5446
jchambers@healthtech.com

CIRCULATION & MARKETING MANAGER
JoAnne Sperry (781) 972-1350
jsperry@healthtech.com

PROJECT/MARKETING MANAGER
Lynn Cloonan (781) 972-1352
lcloonan@healthtech.com

ADVERTISING OPERATIONS COORDINATOR
Matt Gross (781) 972-5481
mgross@healthtech.com

PRODUCTION MANAGER
Tom Norton (781) 972-5440
tnorton@healthtech.com

Contributing Editors
Michael Goldman, Karen Hopkin, Deborah Janssen, Pauline Parry, Salvatore Salamone, Tracy Smith Schmidt, Mark D. Uehling

Advisory Board
Jeffrey Augen, Mark Boguski, Steve Dickman, Kenneth Getz, Jim Golden, Andrew Hopkins, Caroline Kovac, Mark Murcko, John Reynders, Bernard P. Wess Jr.

Cambridge Healthtech Institute

PRESIDENT
Phillips Kuhl

Contact Information
editor@healthtech.com

250 First Avenue, Suite 300
Needham, MA 02494

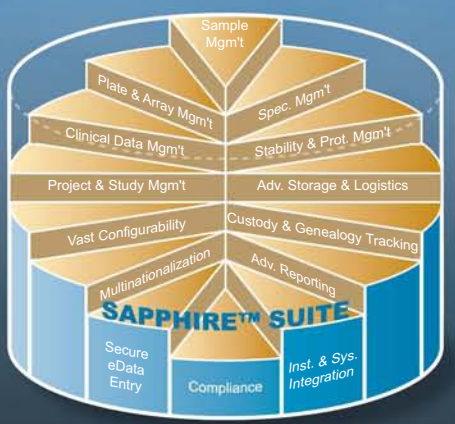
DISCOVERY · DEVELOPMENT · BIOBANKING · FORMULATION · QUALITY MANAGEMENT

LABVANTAGE UNVEILS SAPPHIRE™ R5

A WORLD OF LABORATORY INFORMATION MANAGEMENT OUT OF THE BOX

Sapphire™ Suite: The Best Combination of Functionality, Flexibility and Technology

Successful organizations can no longer depend on one-dimensional systems. As your business grows, so must your systems and their functionality. LabVantage redefines the "S" in LIMS with the ***Sapphire Suite*** -- a zero-footprint and configurable-off-the-shelf Laboratory Information Management *Suite*. Sapphire takes you beyond the world of traditional LIMS by consolidating additional functionality typically found in multiple laboratory systems into one product platform across the global enterprise. Sapphire is the best combination of functionality, flexibility, & technology in one solution.



Configure your success with the Sapphire™ Suite at:

www.labvantage.com/suite

Keyword: OutOfBoxLIMS

IBM Business Partner

LabVantage is an IBM® Business Partner, who has demonstrated success in delivering thin-client laboratory information management solutions for biosciences and pharmaceutical customers. Labvantage's Sapphire runs on IBM on demand software and hardware infrastructure including Linux® software platform and System x. This proven technology is tailored to address the business and IT needs of research, development, formulation, and quality management laboratories. To find out how you can leverage IBM's on demand technology for success in laboratory information management, visit: www.ibm.com/lifesciences.

LabVantage

AUSTRIA • BELGIUM • CANADA • CHINA • DENMARK • ESTONIA • FINLAND • FRANCE • GERMANY
HONG KONG • INDIA • JORDAN • LATVIA • LITHUANIA • LUXEMBOURG • MALAYSIA • NETHERLANDS
NORWAY • OMAN • QATAR • SAUDI ARABIA • SINGAPORE • SOUTH AFRICA • SWEDEN • SWITZERLAND
TAIWAN • THAILAND • UNITED ARAB EMIRATES • UNITED KINGDOM • UNITED STATES

Broad Institute Secures Future with $400 Million Endowment

Broad family announces new $400 million endowment.

BY KEVIN DAVIES

CAMBRIDGE – Billionaire philanthropist Eli Broad and his wife Edythe have announced a new $400 million endowment that secures the future of the Broad Institute.

The announcement was made on the institute's fourth anniversary by Eli Broad in the packed lobby of the building. "This country has been more than good to us," he said. "It's a privilege and a responsibility to give back some of our financial resources to make the world a better place."

Broad said his gift would ensure that the work would continue, making the institute "the place where the greatest discoveries take place." He hoped that his total gifts of $600 million would rise to the $1 billion mark, enabling Broad researchers to continue making headway in the fight to improve human health. To which Broad Institute founding director Eric Lander could just say, "Wow!"

Nobel laureate David Baltimore, the founding director of the Whitehead Institute, who first introduced Broad and Lander, called this "a seminal day in the history of biomedical research." He stressed how essential private philanthropy was in the biomedical research arena. "Why couldn't Eric turn to the NIH [and its nearly $30 billion budget]?" he asked. Because the NIH is "not in a position or of a mindset to be innovative," he said, at least until a new director is appointed.

Among others paying tribute to the Broads for their vision and unsurpassed generosity were Massachusetts Governor Deval Patrick, MIT President Susan Hockfield, and Harvard University President Drew Faust.

Before the announcement, a video tribute to the culture and accomplishments of the Broad Institute featured vignettes from faculty and staff, as well as former President Bill Clinton, Senator John Kerry, and Boston Mayor Thomas Menino. One institute member summed up the collective mood of the "Broadies" saying, "If it's not impossible, we're really not that interested."

ERIK JACOBS/BOSTON GLOBE/LANDOV

Massachusetts governor Deval Patrick thanks Eli Broad at the ceremony announcing Broad's $400 million gift to the Broad Institute.

Lander told Broad he had also received letters of appreciation from Senator Ted Kennedy, Bill Gates, and a certain senator from Illinois with presidential aspirations.

A Bright Future

Eli Broad praised the staff of the institute, saying he was betting on Lander and his team to come up with new ideas, creative approaches, and innovation. He also confessed to knowing little about science or medicine. Much of his philanthropic activity since retiring from commerce 10 years ago has been in improving K-12 education and the arts.

In 2001, Eli Broad gave a grant to Lander for a study on inflammatory bowel disease. The next year, he vested $100 million to establish the institute, which was matched and jointly administered by MIT and Harvard University. The Broads subsequently doubled their donation. The new $400 million investment allows the institute to create a significant endowment that Lander hopes will eventually climb to $1 billion.

Broad declared the past four years "a great success," noting the institute had grown to 1200 people, 330,000 square feet, a grant funding rate of greater than 40 percent, 300 published manuscripts, and a string of notable scientific firsts—20 mammalian genomes, the HapMap project, the Cancer Genome Atlas, and more. Through faculty members such as chemical biologist Stuart Schreiber, oncologist Todd Golub, geneticist David Altshuler, and CIO Jill Mesirov, the Broad has rapidly established itself as a multifaceted post-genomic institute devoted to developing new tools and insights into what is loosely called genomic medicine. •

Merrimack Compound Enters Humans

Their first systems biology product enters human clinical development.

BY JOHN RUSSELL

Merrimack Pharmaceuticals says it has advanced the first systems biology product to human clinical development. Merrimack announced the first patient has received an initial dose in a Phase 1 clinical study of its first oncology product, MM-121, a human monoclonal antibody and a first-in-class therapeutic designed to block signaling of the *ErbB3* receptor.

"We used a combination of high-throughput biology and computer simulation to discern the importance of targeting *ErbB3* signaling in tumor cells," said Ulrik Nielsen, senior VP research. "We believe this is the first systems-designed therapeutic in clinical development. Until now, computer simulation has not been widely applied toward understanding optimal therapeutic strategies for treating cancers driven by complex signaling pathways."

ErbB3 is a receptor in the *ErbB* family, a pathway that plays a critical role in cancer signaling. With the initiation of the Phase 1 trial, MM-121 becomes the first systems biology product as well as the first selective *ErbB3* antagonist to enter human clinical development.

Preclinical data demonstrating the impact of MM-121 in multiple cancer models were presented at the annual meeting of the American Association for Cancer Research in April. The Phase 1 dose escalation study will evaluate the safety and pharmacokinetics (PK) of MM-121. Enrollment is underway at Fox Chase Cancer Center and two additional leading oncology sites are expected to participate in the trial later this year.

FURNALD/GRAY

Ulrik B. Nielsen

ExonHit Bets on Sapio LIMS

Commercial LIMS has the flexibility for the job.

BY JOHN RUSSELL

What Heather Jordan found in her search for a LIMS for biotech ExonHit was a nuanced marketplace that fulfilled and defied its stereotype. In the end, the director of operations for ExonHit Therapeutics was able to find a solid commercial offering—Exemplar from Sapio Sciences—whose key strengths were flexibility, ease of use, and fast deployment.

Heather Jordan

"We evaluated 8 to 10 different vendors," says Jordan. The chief challenge was to find a LIMS suitable for ExonHit's growing microarray processing work and which would be compliant with 21 CFR Part 11 and other regulatory requirements.

ExonHit is a small Paris-based biotech founded in 1997; the U.S. staff is of roughly 20 and there are 50 in Paris. Its proprietary technology—DATAS (Differential Analysis of Transcripts of Alternative Splicing)—monitors alternative RNA splicing events. ExonHit developed a novel design of probes to monitor the differences of highly similar transcripts, received patents on this probe configuration, and in addition to independent research currently runs a service operation that provides researchers with analyzed expression data on individual transcripts.

Regulatory compliance led the list of Jordan's LIMS criteria, as the company is very focused on achieving Good Laboratory Practice (GLP) status. With LIMS users in both the U.S. and Paris offices, she also wanted a single LIMS accessible to folks on either side of the Atlantic.

"We didn't want to worry about installing two different software packages. We just wanted one server that we could both access and not have issues with that. We wanted sample tracking, bar coding, reagent inventory, all of the items that go along with the electronic lab notebook, all the protocol tracking, from sample inventory, and quality assessment to QC [quality control] data," Jordan says.

The LIMS also needed to be flexible. "The majority of our clients at this point are academic labs... We have a portal on our website where researchers can go and select their favorite gene and get a splicing microarray designed for them. These customers have smaller projects, 2, 4, 8, 20 arrays. [But] we also have some pharma with large projects, hundreds of arrays."

The result is the need to track projects that can have either hundreds of samples or projects that have two to four samples, "so we needed something that was very flexible. We needed something that was capable of not only tracking the samples that came in... but we needed to be able to ensure that we could track [samples] from receipt all the way through an array hybridization experiment."

Compatibility with ExonHit's workflow and speedy deployment were also important. "Exemplar had a lot of array protocols and procedures already programmed in for both Affy and Agilent. It allowed for import of the data files... We were really happy we could hit the ground running," she says.

In the end Jordan found two offerings that were suitable, but preferred Exemplar for its functionality and proximity (Sapio Sciences is roughly an hour away). Deployment took a few weeks, says Jordan, and was surprisingly painless for such a big piece of software. The main challenge turned out to be something she hadn't anticipated: rethinking workflows to take advantage of the LIMS.

"I would say we're saving maybe 40 percent of time involved in non-experimental processes. A lot of my time used to be spent writing reports and taking the output files and sorting them for people, burning them on DVDs, all that kind of stuff. Now I can set up a custom report and just click a button and it consolidates all the data and generates a report for me," she says. •

Briefs

PETAFLOP POWER

NVIDIA GPUs are contributing over 1 petaflop of processing power to **Stanford University**'s Folding@home distributed computing network, nearly half of the processing power of the entire project. NVIDIA's contribution is delivered by 11,370 of the total active processors used in the project. Folding@home is Stanford's distributed computing program used to simulate protein folding. Protein-folding simulations can be done 140 times faster on an NVIDIA GPU than on traditional CPUs.

WEB-ENABLED EDC

Phase Forward acquired Radnor, PA-based **Clarix** for $40 million cash. Clarix provides Web-integrated interactive response technology (IRT) used for subject randomization, operational management, and reporting in clinical trials. Phase Forward plans to integrate Clarix' Web-integrated features into its InForm electronic data capture tool. Clarix' platform allows customers to set up clinical trials in half the time of tradition IRT providers.

ANTIBODY ATLAS

Sigma-Aldrich has announced the addition of 2,000 new, highly validated antibodies to the Human Protein Atlas, a non-commercial endeavor of the Human Proteome Resource. Sigma-Aldrich exclusively provides these antibodies to researchers in its Prestige Antibodies library. The Atlas now includes 3,800 antibodies directed against 3,600 genes.

GERMAN ACQUISITION

Bayer has acquired Germany's **Direvo Biotech** for $299 million. Direvo specializes in protein engineering, and will give Bayer a boost in manufacturing biological drugs, such as antibodies, that would presumably add to Bayer's drug unit, which last month won the first approval of Xarelto.

BMS Seeks to Become 'Next-Generation Pharma'

Brisol sees the future of biotech and pharma one pearl at a time.

BY ALLISON PROFFITT

BOSTON—Adopting a more agile biotech philosophy and an acquisition strategy akin to creating "a string of pearls" are the key ingredients to helping Bristol-Myers Squibb succeed as a next-generation pharma, according to CSO and president of R&D, Elliott Sigal.

Sigal, who keynoted DDT in August, believes the industry's current challenges are important, but surmountable. Sigal graduated from medical school in 1981 and began a career spanning medicine, research, and management. Before joining BMS in 1997, he held positions at pharmaceutical company Syntex and the genomics firm Mercator Genetics

From this vantage point, he sees the challenges facing biotech clearly: patent expiration, greater clinical and regulatory requirements, increased role of payers, and "a severe restriction of access to capital." But he also sees much opportunity at this juncture.

"I'm speaking primarily from the standpoint of a mid-size pharma company," he says. "We wonder whether the traditionally do-it-all, own-it-all from every part of the value chain is sustainable. It's certainly a risky and costly model."

Sigal proposes taking what we've learned over the past ten years about biotech and pharma and combining that knowledge into the "next generation biopharma."

"The great attraction of biotechs is their spirit, flexibility, adaptability, and discovery," he says. "You can follow the science essentially wherever it leads and we must continue to support that, even if it means that medicines are ahead but far ahead." Sigal also believes that biotechs are getting it right when it comes to collaborations and partnerships. "I think biotech has really done that extremely well, whereas pharma has just come around to innovative-type partnerships."

"For the best of pharma," Sigal says, "we generally mean selecting and retaining the range of expertise and scope of resources and capabilities that have been part of the industry's heritage including strong small molecule expertise, late stage clinical development, and regulatory expertise. They typically have a base of real financial strength, scale, global reach, commercial expertise, and can work across multiple therapeutic areas and geographies."

Elliott Sigal relies on his medical background to keep the focus at Bristol-Myers Squibb on the patients.

The Best of Both Worlds

The combination of these traits will lead to a next generation biopharma with an innovative portfolio, a continued focus on improving productivity, and a selective integration strategy when it comes to partnerships.

"Part of our selective integration certainly applies to what's been called the String of Pearls strategy," explains Sigal, describing recent BMS acquisitions and agreements including the acquisitions of biotechs Adnexus, and Kosan Biosciences, and licensing agreement with for cardiovascular disease with Kai Pharmaceuticals. Interestingly, Jeremy Levin, formerly of Novartis, has recently been named head of BMS' new Strategic Transactions Group. (See, "Leading by Example," *Bio•IT World*, March 2007)

"The goal of these transactions is consistent to access innovation across specialty and biotech wherever it may be, or in pharma, that will enhance or complement our own internal expertise and build strength in key areas." (As "what we hope will be a fourth addition to the string of pearls," Sigal acknowledged BMS' bid for ImClone, but only to read a brief statement saying the company believes the transaction to be a strategic next step of ImClone's existing relationship with BMS.)

This is all very strategic, Sigal says. "We actually set our discovery engine to be less capacity than what we thought the whole organization needed so that we would force our portfolio strategy to go outside, level attrition, and to access innovation."

With their identity as a next generation biopharma in mind, Sigal said that Brisol-Myers Squibb is focusing on choosing "disease areas where we saw significantly unmet needs and had the core competencies to address R&D and commercialization," "multiple modalities to find the right approach to the right target," "continuously pushing ourselves to develop medicines in different ways," and developing new treatment options that "demonstrate compelling economic value for delivering that medicine to patients."

He acknowledged that his plan isn't for everyone. "I think for some companies this is a map for the next generation biopharma, not for all," Sigal said. But he believes BMS is perfectly positioned. "Major biotechs and mid-sized pharma, I believe, are in the best position to have the agility and capabilities to make that bridge." •

* IBC's Drug Discovery and Development of Innovative Therapeutics, August 4-7, 2008, Boston

Litigating Patent Obviousness

DAVID FRAZIER and WILLIAM RAICH

Sound patent protection is a necessity to justify the time and expense of developing new pharmaceutical products. To be patentable, an invention must be completely novel and be the result of inventive activity, that is, it would not have been obvious to a skilled person at the time it was made. But when is this the case?

In spring 2007, the Supreme Court ruled on *KSR International v. Teleflex*, commenting for the first time in decades on the meaning of "obviousness" under patent law. Although *KSR* did not involve a pharmaceutical patent, there has been concern that the ruling would make it harder to defend existing patents. Moreover, some infringers argued that principles of chemical patent law should be jettisoned to the detriment of pharmaceutical patentees. Fortunately, as the dust clears, the *KSR* case and subsequent decisions are providing a clearer map for navigating this uncertain legal landscape.

In *KSR*, the Supreme Court struck down as obvious a patent for an adjustable automotive pedal assembly, declaring that the pedal assembly merely combined previously known parts (such as an electric sensor and an adjustable pedal) to yield "predictable" results. According to the Court, when a patent simply combines previously known elements to produce a result that was no more than expected, the combination would be obvious. On the other hand, inventions involving known elements that work together in an unpredictable manner would not be obvious.

Thus, an important principle reaffirmed in *KSR* is that a patent is more likely to be upheld if an invention is *unpredictable* in view of what was already known. This is good news for pharmaceutical innovators, because drug development is notoriously unpredictable.

Unpredictable Benefits

The Supreme Court in *KSR* provided guidance on predictability: if there is market pressure to solve a particular problem, few possible solutions, and predictable results, then a skilled person has good reason to pursue the invention, and one's ultimate success might be viewed as obvious. However, if the field prior to the invention is littered with cautionary tales that "taught away" from the invention—such as failures or concerns about the feasibility, safety, or efficacy of the invention—then the inventor's ultimate success would not be deemed obvious.

This emphasis on unpredictability was illustrated in recent litigation concerning Eisai's antiulcer drug AcipHex (rabeprazole), which garners over $1-billion in annual worldwide sales. The infringer alleged that the patented compound was obvious because a skilled person would have known to prune a particular chemical side chain from a structurally similar known compound to yield the patented drug. The court, however, determined that this modification was not a "predictable solution," because that side group was credited in a prior publication with bestowing certain desirable traits. Consequently, the advantages of pruning the side chain were not predictable and the patent not obvious.

Litigation involving Takeda's Actos (pioglitazone), a $2-billion diabetes drug, demonstrates the importance of "teaching away" to assessing patent obviousness. The infringer alleged that the patented compound was obvious over a known "compound b." However, compound b had negative effects, including increased body weight. Because of these effects, the court determined that modifying compound b was not a "predictable" solution for diabetes.

What does all this mean for pharmaceutical patent owners? While it is relatively common to document surprising or unexpectedly positive results, it may be similarly helpful to catalog the blind alleys and failures that predated ultimate success. We recommend close coordination between research and legal functions to identify and maintain helpful documents and to interview key individuals. Even if it is not possible to follow such an approach with every patent filing, if you are (or anticipate) defending a patent's validity, it is never too early to begin compiling evidence of unpredictability and teaching away. Commercial pharmaceuticals are typically the lucky result of extensive screening and testing from laboratory assays to clinical trials. Documenting and telling the story of how a particular compound beat the odds can be the difference between success and failure in patent litigation.

A patent is more likely to be upheld if an invention is unpredictable.

The *KSR* decision weakened predictable mechanical combination patents by making them easier to invalidate. But the unpredictability of drug development suggests that *KSR* will have less effect in pharmaceutical patent litigation. Innovator companies can inoculate themselves against patent challenges by clearly establishing actual facts that taught away from their inventions.

David Frazier and William Raich are patent attorneys with Finnegan, Henderson, Farabow, Garrett & Dunner, LLP. They can be reached at david.frazier@finnegan.com and william.raich@finnegan.com.

Computational Biology

Febit Puts Customization to the Test

Launch of RT Analyzer offers bespoke biochips, expanding applications.

The Geniom RT Analyzer uses custom chips that can be synthesized in the febit store.

BY KEVIN DAVIES

Michelle Lyles has a flying pig in her office. It's a long story—something to do with her former employer and the challenges of raising fat sums of venture capital. She hopes it's a good omen for her latest challenge at febit, which has reached a critical crossroads in its efforts to wrest business away from the titans of the microarray field—Illumina, Affymetrix, Agilent, and Nimblegen. Febit hopes that the commercial release last month of the Geniom RT (RealTime) Analyzer, a $100,000 benchtop instrument, will provide the flexibility and affordability researchers are seeking.

Febit was launched in Heidelberg, Germany, in the late 1990s. Of the name, "Fe" is derived from the last name of the co-founders, Cord and Peer Staehler (meaning 'steel' in German), while the bit is for bits and bytes. The majority shareholder is Dietmar Hopp, who co-founded SAP and is a major stakeholder in German biotechnology.

Last month's launch of the RT Analyzer is a strategic shift away from the company's flagship instrument, the Geniom One, which both synthesizes the biochips and serves as the experimental platform. "The mission is the same but the commercialization path has changed, because the market changed," says Lyles, VP sales and marketing. "By the time we got out there with the Geniom product, you had the Affys and Illuminas and Agilents that provided the commercial high-density chips."

14 Geniom One instruments have been sold since 2006. It is expensive (around $275,000) and has a lower probe density than competitors that provide high density chips. The RT Analyzer provides the same core array technology, but now febit is handling the rapid custom synthesis.

Three Pillars

Lyles says febit has "three pillars of technology"—engineering, bioinformatics, and applications. The engineering strength is in the microfluidics, integrated with temperature control and digital micromirror device (DMD) technology from Texas Instruments. "It's the same DLP chip you find in a projector," says Anthony Caruso, VP global informatics and former president of LION Bioscience. "When febit first started, they'd buy projectors and take the DLP chips out of the machine, because we were too small to get them from directly Texas Instruments."

The integration of microfluidics and DMD technology provides rapid, highly customizable synthesis of biochips. The mirrors shine certain wavelengths onto the light-sensitive protector groups at the tip of each oligo probe, enabling elongation to occur. The array is not a single glass slide but a sandwich, providing a 3-D tunnel with eight microchannels and temperature control.

The other two pillars are bioinformatics and applications, which include HybSelect (see, "Hope for HybSelect") for resequencing on next-gen sequencing platforms, synthetic biology, and microRNA.

Febit is probably not the platform for those contemplating large-scale, genome-wide association studies. "We thought about competing with Affy [initially], but given that they have ten times as many features, we made the strategic decision to turn our focus to targeted analysis and enhanced applications" says Caurso. "We can't do as many features, but you don't need to do as many features if you're doing targeted analysis and new applications."

Theoretically, febit could design up to 500,000 SNPs on a biochip by utilizing every mirror. But Caruso points out the Geniom CCD camera does not offer the same resolution as a scanner, so he prefers to leave a space between the features. The highest practical density for expression profiling is about 250,000 features in a checkerboard pattern, all of it completely programmable on the instrument. Using every feature might be feasible in HybSelect and other applications that do not require imaging.

Adds Lyles: "Because we're using CCD imaging for detection rather than fluorescence scanning (which employs a laser), you can read the chip over and over again, so you get data in real time. This enables dynamic detection and opens the door to new applications for microarrays including melting curves, and thermocycling, which you can't do with other technologies." Researchers can do PCR amplification and primer extension on

the chip, which enables applications such as FFPE (formalin-fixed paraffin embedded) analysis. "You can't do that on an Affy or an Illumina," she says.

Custom Ordering

Febit has also found that many researchers preferred not to synthesize the oligo probes themselves when they had the convenience of ordering bespoke biochips. "We'll synthesize the chip for you and send it to you," says Caruso. "Then you use the RT Analyzer for the experiment." Alternatively, researchers can use febit's experimental service.

"It's like the Apple store," he says. "If you have the RT Analyzer and you can't synthesize the chip, you still have all the tools available. You can go to the febit store—you can pick the genes you're interested in, and then electronically—once you've designed what you want—you send it to the febit store. We get the order, synthesize the chip, and send it back to you."

For targeted resequencing or HybSelect, febit provides users with software that takes genomic and SNP data and calls the probes in any given chromosomal region. Synthesizing a biochip of "shortmers" (about 25mers) would take a mere six hours—whether in Germany or febit's U.S. facility in Lexington, Mass. Febit promises a two-week turnaround time, though Lyles expects it will be quicker than that. And once the chip has been designed, febit will keep it in stock if requested. "We're encouraging people to run a trial first, including the design of a chip. For a customer that wanted to start an experiment, you could do a trial run using our services," says Lyles.

Another key advantage of the febit chip is flexibility says Caruso: "One of the beauties of the instrument is we can do any organism at this point." A good example is the first draft of the sea urchin genome. "We cherry picked probes for those genes including replicates," says Caruso. "You couldn't go with Affy because they don't make sea urchin chips."

The ease of biochip synthesis affords a convenient shuttling between theoretical bioinformatics calculations and empirical wet lab experiments. These allow several iterations of a particular chip design to allow researchers to run a few cycles during a week or two to develop the optimal assay. "The flexibility doesn't cost us anything extra," adds Lyles. "We don't hold customers hostage to the idea that they

(CONTINUED ON PAGE 16)

Lyles says febit considered competing with Affymetrix, but decided to focus on targeted analysis.

Hope for HybSelect

The first pilot user of febit's HybSelect technology is associate investigator Matthew Huentelman at the Translational Genomics Research Institute (TGen) in Phoenix. Febit plans to officially introduce the technology in early 2009.

Huentelman is an enthusiastic user of the Geniom One, which TGen has used since 2006. "It's exciting because it's fully user programmable," he says. "You can print whatever you want—whatever you dream up today, you can print it and come in the next day and use it in your experiments." He studies the genetics of autism and Alzheimer's disease. "We can tile across an interesting region of the genome with full customizability. You're able to make the chip in the same machine that you use it." The chip design comes into its own "when you're ready to validate some regions. The power is the customizability."

The TGen group might start by identifying copy number variations (CNVs) using a typical SNP array. From there, it uses the Geniom One to fine map the CNV region and identify the relevant genes. The HybSelect protocol selects specific regions for targeted resequencing using next-gen platforms. "The goal is to use this approach to sequence focused regions of the genome across large numbers of individuals," says Huentelman.

One typical experiment might be to take 100 key genes, make a custom array of exons of those genes to enrich those sequences from the rest of the genome. Those genes are captured on the chip, eluted, and then run on a next-gen sequencing platform.

The probe design process is "user friendly as long as you have a good bioinformatics team on board," Huentelman says. "You want to provide intelligently designed probes to ensure the highest quality data. You need to have a standalone oligo picker that picks the probes for you. We prefer to pick the probes ourselves. The output from that gets uploaded."

Huentelman admits, "we're still trying to fully characterize the degree of enrichment we are seeing when using HybSelect." If he has one major criticism, it is the number of features on the chip. "It's great to be customizable but we need a jump in density for them to be competitive." He adds: "The field would really benefit from a robust cost-effective approach for enriching regions of the genome for sequencing—that's the ultimate goal in our pilot use of febit's HybSelect."

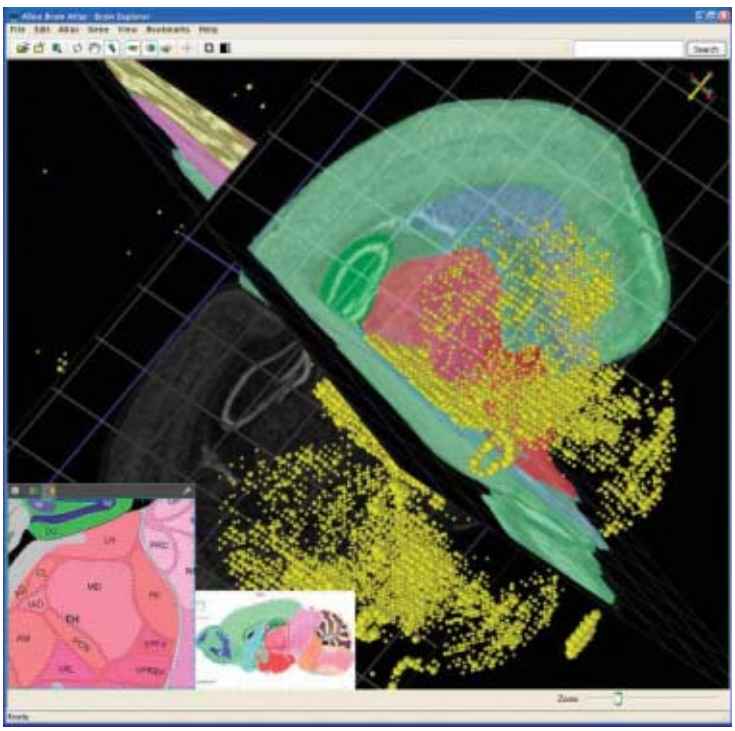
Gene Cartography in the Brain

Allen Institute for Brain Science expands its atlas collection.

BY JIM KLING

Founded in 2003 with $100 million in seed money from Paul Allen, the Seattle-based Allen Institute for Brain Science (AIBS) is a short walk from the center of the well-known art district of Fremont. Institute offices overlook a portion of the canal that links Lake Washington and Lake Union, treating employees to the occasional views of kayaks and other watercraft.

There's not much time for people watching at AIBS, though. The institute created its flagship Allen Brain Atlas (ABA) to fuel neuroscience discovery. Since its initial publication in September 2006, the ABA has been a boon to neuroscientists worldwide. The project was ambitious: it generated a 3-D map of gene expression at cellular resolution in the mouse brain that can be visualized by specific genes and through anatomical reference points. The atlas is freely available online and has been cited more than 150 times, says Kelly Overly, research alliance manager for AIBS.



The Brain Explorer

Last July, AIBS published the first results of a follow-up project—the mouse-based Allen Spinal Cord Atlas. Currently, about 2,000 genes have been mapped, with the institute projecting genome-wide coverage by the end of the year. It too can be searched and sorted by gene, age, expression, and anatomical structure.

Two other projects are farther from completion. The ABA Human Brain will be complete in about four years. The ABA Developing Mouse Brain will follow gene activity across different stages of development between birth and adulthood and will be completed in two years. These projects represent a major expansion for AIBS. To drive them, Overly anticipates a 50% increase in the scientific workforce by the end of 2008.

Cutting Edge Informatics

AIBS's researchers largely use off-the-shelf technology—microscopes, cameras, slide carriers, automated stages—but focus on developing custom software to handle parallel imaging systems. The ABA employed ten systems in parallel running 24/7 with little operator control.

"It was largely a matter of putting things together in a way that hadn't been done before. The challenge, like with any biological system, was to develop algorithms that can be robust but still account for the inherent variability in biology," says Overly. The researchers employed a bright field microscope for greater resolution with 10x lenses. Each image section was broken into one hundred tiles that yielded cellular-level resolution.

The greatest hurdle was the downstream processing, Overly says. The team wanted to encode, align, and package the data in a way that would allow users to visualize it at will, like a virtual microscope. They also wanted users to have good flexibility in mining the data.

The spinal cord project required some modifications to the LIMS system, as well as new methods for sectioning and slide preparation. Moreover, the imaging software had to be taught to recognize the new configurations. But the basic technology is the same. "We're just tweaking the system to account for different-sized tissue specimens and different signal characteristics due to the nature of the tissue," says Overly.

Brain Scans

The net result is a publicly-available web site that includes some powerful tools. The Brain Explorer desktop utility allows users to explore 2-D data sets in a 3-D representation. Researchers can also do searches by anatomical structure. Says Overly: "One of the most powerful tools is the search function 'genes like me.' Say a user is interested in the pattern of expression [of a particular] gene in the hippocampus. You can seed the system with that gene and tell it to show all other genes with similar expression patterns in the hippocampus. That allows the user to pull out groups of genes that might

be [related] to one another by virtue of proximity. Having the whole database of genes and being able to combine spatial information with the gene data gives them a jump start in their research."

Another tool, the Anatomic Gene Expression Atlas, gives an unbiased view of the brain that is driven by gene expression, according to Overly. "It allows you to ignore every classical marker of neuroanatomy—you're just looking at gene expression and seeing which areas are most similar to each other. It shows a (genetic) organization of the brain. What you find is that in many cases the gene expression patterns line up reasonably well with the anatomical divisions that have been established using other methodologies during the years. If you go to higher and higher resolution, gene expression might be able to inform some of the more fine boundaries within the brain. It allows users to understand the organization (of the brain) in a different way."

The anatomical structure of the spinal cord is simpler than that of the brain, so the spinal cord atlas doesn't include a 3-D reference atlas. "It's a bit more streamlined database with images of gene expression," Overly says. ABI plans to release anatomical reference images that can be used with the data.

Mind Control

The two atlases are extremely valuable to neuroscientists, says Jane Roskams, associate professor of neuroscience at the University of British Columbia: "The strength of a tool like that isn't just that it helps you find clues to pathways, but it also tells you which pathways are not important (to your system). It's easy to overlook that because it doesn't end up in a published paper, but it gets you ten steps further (in your research) than you would have been if you didn't have access to that information."

Roskams helped AIBS locate spinal cord experts to advise on the project. She soon realized how little is known about the system. "We don't know the function of many of the cells," she says. "We know the region they're in—regulating pain, or sensory signals, or near motor neurons that drive large cells—but we've only scratched the surface of understanding how they work together."

The developing brain atlas will complement the other gene atlases. Roskams expects it to provide a road map to guide spinal cord regeneration researchers, for example, while the brain and spinal cord atlases can act as a measuring stick for success. "On the one hand [the developing brain atlas reveals] the programs you need to follow to reconnect areas of the spinal cord [following injury], and on [the spinal cord atlas] is a readout of what you should expect. It helps us to know whether or not we've reached the end zone," says Roskams.

The projects aren't the only ones of their kind, but AIBS' resource base and institutional focus puts it in a unique position, Overly says. Other projects do not have the breadth and scope of the AIBS projects. "It's challenging because everyone has different standards, and the volume of data produced is tremendous," he says.

Roskams appreciates the effort. "It's a long and tedious road to get that data and do it well, even for a single set of genes. They're taking the tedium out and giving [the data] back" to the research community, she says. Still, Overly hopes that the research community will play an active role. The projects are designed to be open access. Users can also access the data itself, including the 3-D coordinate system and XML files of gene expression data. "We're working hard to make as much of it open as we can," says Overly. "Users can bring the data in house in a way that they can create mash-ups or mine it and manipulate it, or develop their own technologies on top of it." •

Febit

(CONTINUED FROM PAGE 14)

want to customize. We treat a 'custom design' the same as something off the shelf design."

Niche Role

Febit is one of several vendors that will be featured in an imminent publication from Harvard Medical School scientist Winston Kuo, comparing a handful of array platforms. The study evaluated emerging array and nanoscale volume technologies including ABI Taqman, BioTrove, Fluidigm, Roche 480, NanoString, and Phalanx Biotech. "We are witnessing a trend in the microarray community towards developing and utilizing focused arrays," says Kuo. Full details must await publication, but Caruso says, "Our performance is very respectable."

Among the near-term applications for the RT Analyzer will be targeted enrichment for next-gen resequencing. "You can use HybSelect to reduce the complexity of your genomic DNA sample by targeted hybridization and elution of regions of interest," says Lyles, who anticipates a huge market in the coming years. Unlike other services, "we're the only ones who offer the hardware to do that," says Lyles. Caruso is working on a suite of bioinformatics modules to design chips for the RT Analyzer for targeted sample enrichment. Customers would purchase the module, put in the instrument to enable the Analyzer to do the hybridization and elution.

Another growing area is microRNA profiling. "We have the chip. We're always the first," says Lyles. Four months after releasing miRBase 11.0 biochip, febit is out with version 12.0—one day following release of miRBase 12.0 from the Wellcome Trust Sanger Institute. And there's additional room on the biochip for groups to add their own microRNA targets. Disease chips, pathway chips, and protein-DNA binding round out the offerings.

Lyles admits the market demand to synthesize custom chips is currently lacking, but expects febit's new commercialization path to be ready when it rebounds. "The RT Analyzer will be the workhorse, and once it has a foothold in the market, then let's put the Geniom One in core facilities to feed chips to the satellites."

Febit's U.S. facility is awaiting its first RT Analyzer. The second is promised to Prognosys in San Diego, a company founded by Mark Chee, who previously founded Illumina. The company hopes it will be the first of many. •

Computational Development

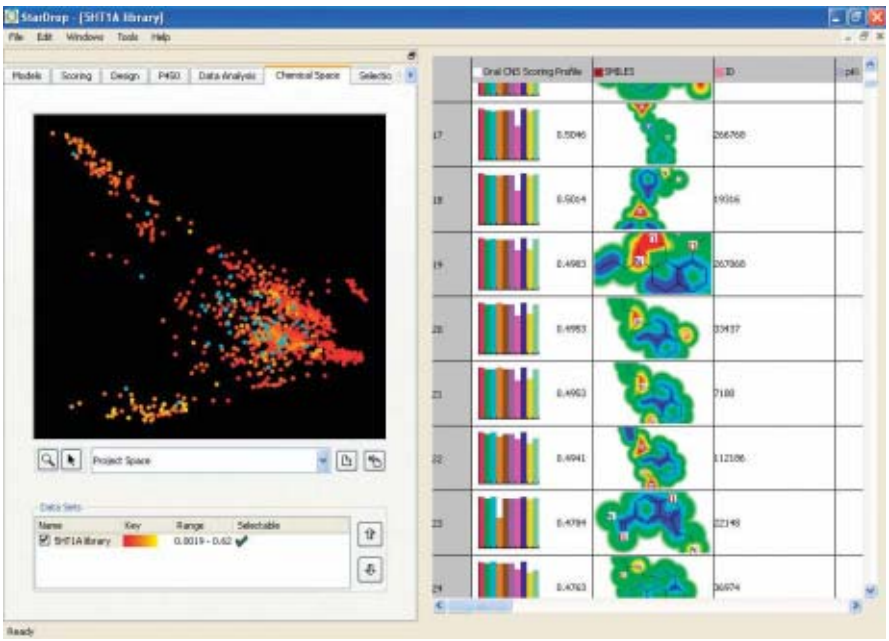
Picture Im-Perfect

By expecting less, pharma can get more out of modeling.

BY VICKI GLASER

Software packages described as decision-enabling or decision-support tools are gaining a higher profile in the drug discovery arena, yet how enabling are these predictive *in silico* approaches? Are medicinal chemists confident enough in their predictive capabilities to rely on them for compound design and lead selection?

The sorts of software packages being marketed as decision support tools typically offer either data integration packages, LIMS systems, or graphic visualization tools, but "do not make a real difference in the way decisions are made in drug discovery," contends Matt Segall, senior director of ADMET at BioFocus DPI. In his view, *in silico* modeling software should enable "guided decision making—it's about analyzing the data within the context of your objectives to identify the best route to take forward. Because of the complexity of the data now generated in drug discovery, that is a multi-component, multi-objective optimization process."



BioFocus DPI's StarDrop software enables drug discovery scientists to design and identify balanced compounds most likely to succeed using all available data and focus resources on the most appropriate chemistries by exploring the full "chemical space" of a project.

Too often overlooked is the fact that "every individual parameter you are measuring has a significant degree of uncertainty or variability, and that needs to be taken into account when making decisions," says Segall. He describes BioFocus' technique of "probabilistic scoring" as a method that applies probability-based analysis to complex, uncertain data to identify areas of chemistry that have the highest likelihood of success.

The company designed its StarDrop platform for use by drug discovery scientists rather than computational experts, to help them define the criteria for success of a project. As key criteria may be in conflict—for example, driving potency up may sacrifice metabolic stability or other pharmacokinetic properties—part of this process involves defining the relative importance of individual properties and generating a target product profile.

The software helps identify and balance these inevitable trade-offs in multiples properties according to individual scientists' priorities in a common and objective language. If those priorities change with future analyses and an expanding knowledge base, then the profile can be modified and the test compounds rescored to determine what effect the change has had on the decision-making process.

Although a target profile is itself subjective, Segall would argue that "it's better to decide up-front what you are looking for—but be flexible—than to focus too much on one objective and hope you end up in an area of chemistry that will give you the other properties you need."

Why the Bad Rap?

Mike Moyer, director of medicinal chemistry at the Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, spent 18 years at Pfizer before moving to academia. In his current environment the more limited resources available to synthesize and test compounds has put a premium on optimizing compound design early on. To achieve that, Moyer's group has turned to *in silico* modeling techniques and has, in a sense, turned the model generation process on its head.

Moyer contends that building the most predictive *in silico* models while synthesizing the fewest number of actual compounds requires an up-front investment in model generation. This necessitates designing and synthesizing compounds specifically for the purpose of constructing and optimizing models.

"*In silico* models will only be as good as the datasets on which they are built," says Moyer. Although he acknowledges that "no medicinal chemist wants to make a compound simply for model making," he believes this to be the most efficient path to compound optimization.

Computational Development

The initial compounds made in many medicinal chemistry labs do not necessarily yield optimal models. Thus, when these models are then used to drive compound design and optimization, the results may be disappointing, which, Moyer states, is at least part of the reason *in silico* modeling may not be as predictive as desired and has garnered a bad reputation.

Unrealistic expectations have also plagued the field. *In silico* modeling programs "are not perfect and they will never be perfect," says Paul Wyatt, director of drug discovery in the College of Life Sciences at the University of Dundee. "In the world of [drug metabolism and pharmacokinetics] (DMPK), there is never an exact figure"—a drug does not behave exactly the same way in two different people.

After a 25-year career in pharma and biotech, including stints at SmithKline Beecham and Glaxo, Wyatt moved to Dundee in 2006 to set up a drug discovery group focused on neglected diseases, mainly tropical parasitic diseases. Based on his experience he is convinced that "*in silico* must be the way to go for predicting the properties and activities of molecules."

The value of *in silico* modeling, asserts Moyer, lies in the ability to perform multidimensional computational analysis and to strike a balance between competing properties and identify the best possible combination of features of an individual compound designed for a specific target.

Moyer values the StarDrop program for its ability to look at the trade-offs between potency and ADME properties and "find ways to get to the sweet spot" as quickly as possible. "It has a way of explicitly calculating the uncertainty in the model and making that visible to the user. It gives you a value along with an associated error and makes clear the limitations in the model."

> *In silico* modeling software should enable "guided decision making—it's about analyzing the data within the context of your objectives to identify the best route to take forward."
>
> **Matt Segall**, BioFocus DPI

Defining the Anti-Drug

Perhaps the greater benefit of *in silico* modeling lies not in identifying promising compounds, but rather in ruling out compound sets or chemistries that would likely lead down a development path doomed to failure.

"These programs can give you a pretty good idea of which compounds not to make," says Wyatt. Pharma has been operating under the faulty assumption that if it takes 10,000 compounds to find a drug, then there must be a suitable drug candidate in a library of 10,000 or more compounds, and it is just a matter of making the compounds and screening them against the target to find the drug. But actually synthesizing large compound libraries eats up valuable time and resources and is no guarantee of success.

One of the limitations Wyatt has found with many of the modeling and decision support tools available on the market is that they are standalone programs that do not allow for communication and data sharing between applications and users. With StarDrop, Wyatt's group is able to incorporate its own data and scoring functions, to compare sets of compounds, and to generate compound profiles across a broad range of parameters.

StarDrop incorporates *in silico* ADME and QSAR models, quantum mechanical P450 models, probabilistic scoring techniques, data analysis tools, the Glowing Molecule visualization and interactive compound design tools, and the Auto-Modeler for developing models based on users' own data. Earlier this year, BioFocus DPI made several of its predictive drug discovery databases publicly available online by transferring them to EMBL's European Bioinformatics Institute. The StarDrop software platform was not part of this transaction, which did include DrugStore, StARLITe, Strudle, and Kinase and GPCR SARfari. In August, Galapagos acquired Sareum Holdings' drug discovery services business, adding structural biology capabilities to BioFocus' pre-clinical drug discovery services. •

Bio·IT World

Best Practices AWARD

2009

CALL FOR ENTRIES

The Bio-IT World Best Practices Awards 2009

"[The Best Practices Awards] is an indication of how we are all working together in this era and that is a wonderful thing."

–Dr. Francis Collins
Natl Hum Gen Res Inst, NIH

"Winning the award increased our visibility to the external community, especially with regard to the rigor with which we do our science. That, in turn, enhanced our ability to get funding and hire people to continue on our growth trajectory."

–Dietrich Stephan, Ph.D.
Co-founder & CSO, Navigenics; former Deputy Director, Translational Genomics Research Institute

Bio-IT World is accepting entries for its prestigious 2009 Best Practices Awards program. Established in 2003, the program recognizes teams for their novel and innovative uses of technology, business strategies, and solutions benefitting the biosciences value chain, from basic research to clinical trials. Direct entries are encouraged as well as nominations from users and vendors.

All entries are reviewed and the winners selected by a distinguished panel of experts from industry and academia. Awards are presented at a gala dinner held on April 28, 2009, in conjunction with the 2009 Bio-IT World Conference & Expo, World Trade Center in Boston.

Highlights of the 2008 Best Practices Awards Program

"Bio-IT World salutes the winners of its 2008 Best Practices Awards program. Not only have they demonstrated impressive technical achievements as judged by their peers, but also they've demonstrated a commitment to advancing "best practices" throughout the biomedical research and drug development community by their willingness to share information about the innovative tools they are using and explaining how they use them."

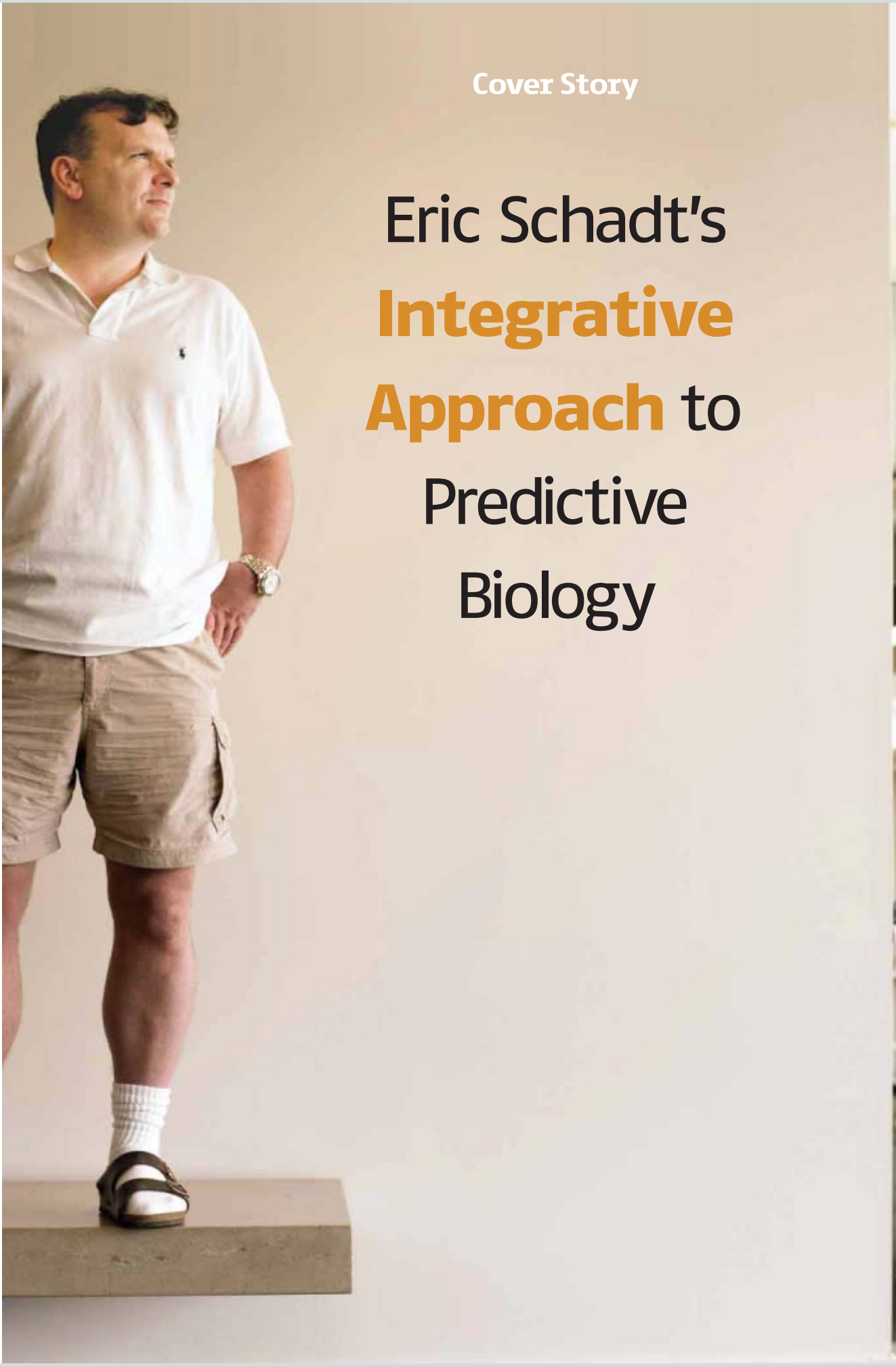
– Kevin Davies, editor-in-chief of Bio-IT World

Submit your entry today!

Visit www.bio-itworld.com/bestpractices for entry details and submission forms.

Cover Story

Eric Schadt's Integrative Approach to Predictive Biology



The difference is in the layers. Eric Schadt believes that genome-wide association studies are just the beginning.

2007 was proclaimed as the *annus mirabilis* for genome-wide association studies (GWAS)—a tipping point where scientists were able to identify scores of susceptibility genes for common diseases. But scientists at Seattle's Rosetta Genomics, a subsidiary of Merck, led by Eric Schadt, are aiming for something altogether grander.

Schadt's assertion that GWAS "doesn't really take you very far," is bordering on heresy these days. His view, however, is that while an association study provides a telling signpost in the genome for a specific trait, it doesn't necessarily pinpoint the causative gene to treat that disease. "Even if it does lock you onto the [susceptibility] gene, it doesn't tell you: do I activate the gene? Do I inactivate the gene? And it doesn't tell you what is the broader context in which that gene's operating in? What's the relevant tissue? And how is it actually leading to this disease state or drug response phenotype?"

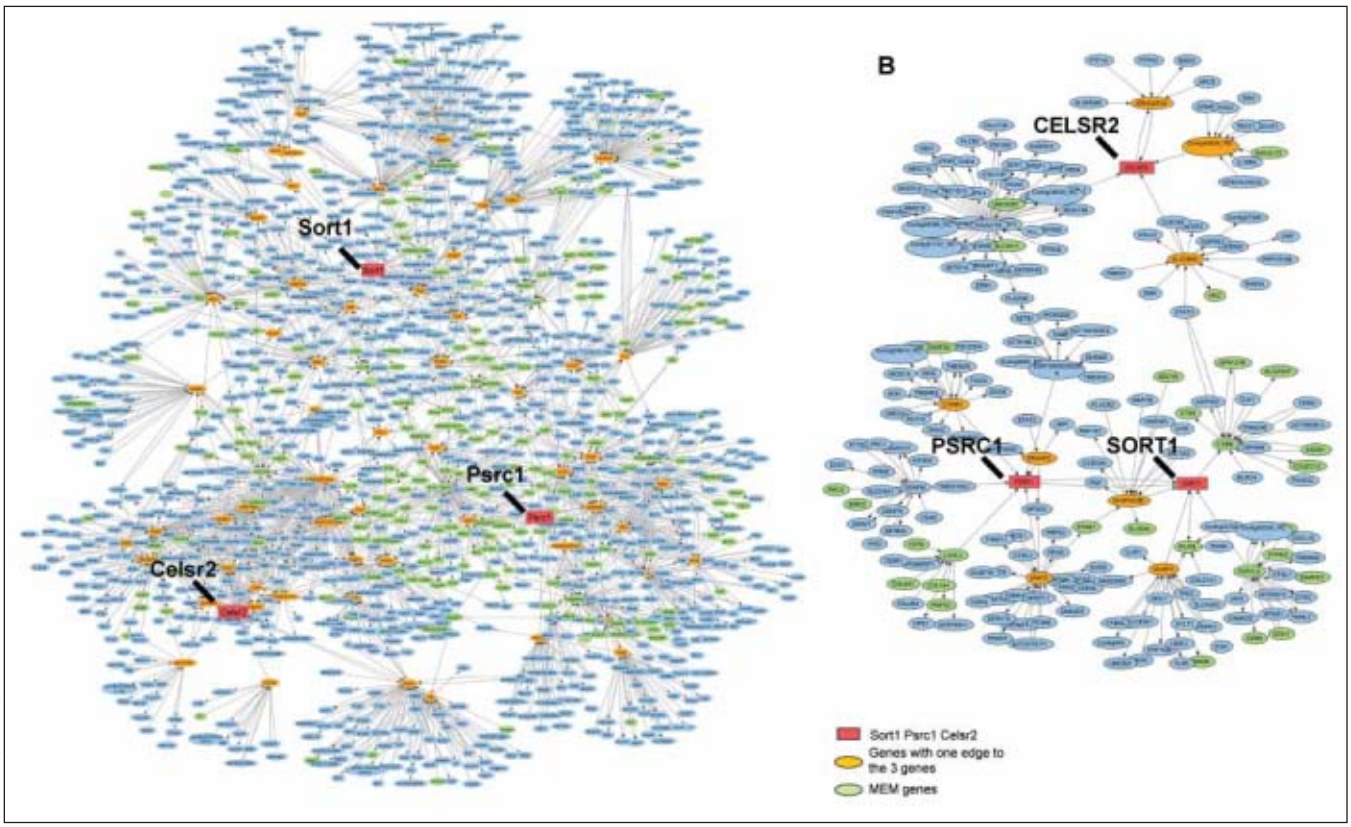
And so Schadt, a mathematician by training who joined Rosetta Inpharmatics in 1999 partly out of frustration with the limitations of academic grant funding, has built an integrative genomics research program that begins by identifying key disease-related genes and then asks: What are the networks involved in driving those disease states?

Schadt sees genetics as simply one dimension of a much more complex problem. "Changes in DNA don't directly cause obesity or diabetes. Instead they're affecting these molecular networks—gene networks or molecular networks or protein function or transcription or metabolite concentration, whatever the case may be... It's changes in those molecular states that go on to predispose you or increase your disease risk."

If Schadt's group can "fill in those middle layers, those intermediate phenotypes that respond more proximally to the DNA variation," then it can reconstruct a more accurate picture of how disease manifests itself. And Merck, which owns Rosetta, is happy. "We can deliver completely novel, unexpected, well-characterized targets that nobody else is working [on]," says Schadt.

By John Russell & Kevin Davies

ANDY REYNOLDS



PLOS BIOLOGY, MAY 2008, VOL. 6

Combining mouse and human data, Schadt's group identified three genes—*CELSR2*, *PSRC1*, and *SORT1*—that seem to operate in a conserved subnetwork causally associated with cholesterol levels, obesity, diabetes, and atherosclerosis.

Schadt's rapidly rising profile has certainly caught the field's attention. Incoming requests are "completely out of control," including up to 200 emails a day. He receives 100 lecture invitations each year. Even rival pharmas such as Pfizer and Amgen are asking him to give talks to their researchers.

IT PAYS TO LAYER

The idea, which Schadt has been nurturing since publishing a major paper with Merck vice president Stephen Friend in *Nature* in 2003, is to consider DNA variation information as but one dimension, and layer on top data on gene expression or protein interactions or DNA-protein binding. Then integrate that information into a mathematical model to construct networks that are predictive of these disease states.

Once you have a predictive model of a disease state, Schadt can say: If we have all the genes involved in this network, what's the best node in the network to target disease treatment or drug response? "It solves a lot of the biology upstream, so that you can make better informed decisions earlier on about what you should be targeting, what you should be tracking to stratify patient populations into disease subtypes and responder groups... so you can make more informed decisions."

Five years ago, Schadt's group started a collaboration with Iceland's deCODE Genetics, culminating in one of two landmark papers (this study looked at blood and adipose tissue in the Icelandic population) that appeared back to back in *Nature* in March 2008. There are about eight different projects ongoing in different tissues and diseases, including brain, cancer, and the liver (see below). Each one takes roughly a year to complete.

WELLCOME RELIEF

In a recent report in *PLoS Biology*, Schadt's group looked at the "amazing" Wellcome Trust Case Control Consortium GWAS study from 2007 (*Nature*, 07 June 2007), which mapped genes for seven common diseases in 14,000 individuals.

In type 1 diabetes, the most significant association had come with the gene *ERBB3*. But Schadt noted that the only evidence was that the diabetes-associated SNP was situated closest to *ERBB3*—but it wasn't actually in the gene region. Although from a biological standpoint *ERBB3* made sense, Schadt's team set out to examine all neighborhood genes and objectively monitor their activity in tissues relevant to diseases like obesity and diabetes.

When the team looked at gene expression levels across 400 samples, asking whether the activity of any of the 50-100 genes in that region correlated with the SNP association, *ERBB3* struck out. But a nearby gene, *RPS26*, was significantly associated. 40% of the variation in *RPS26* was explained by this SNP. Says Schadt:

"Now you have this SNP that's associated with disease that's also associated with the activity of this gene that's right in the region."

But even that's not enough. Next, Schadt's group reconstructed the information in a network of tens of thousands of genes. He asked: what's the context in which this gene operates and does that context support type 1 diabetes? "This is the power of having constructed these predictive models—we can go into that network and say, give me the location in the network where *RPS26* resides and then give me all the genes in the region of *RPS26* that *RPS26* is communicating with."

Schadt combines these data into a causal probabilistic network, which predicts not only what genes are interacting or correlated with each other, but which one is actually causing others to change. The DNA variation information is critical because it is the ultimate source of perturbation. Once the network is constructed, it is represented in a database that can be searched *in silico*.

By querying all the genes the network predicts are directly affected by *RPS26* or directly affecting *RPS26*, the search produces a subset of genes. Then the team asks, are the genes representative of type 1 diabetes? In a query against the KEGG database, the type 1 diabetes pathway was the number one pathway enriched in this RPS26 sub network.

"That's just an example of how we can show how this approach informs the discoveries coming out of the GWAS." The integrative genomics approach provides information about what is going on in that region; what is the gene, what's the network it's operating in, and a functional mechanistic understanding of its effect.

DATA DIVERSITY

According to Schadt's new study in *Nature Genetics*, increasing molecular state information leads to more predictive models. That means scoring not just DNA variation or gene expression, but protein interaction, DNA protein binding, differential methylation, and so on.

Schadt says the tools for detecting DNA protein binding and methylation status are definitely improving. "Within the next few years, you'll start seeing large-scale, population-based studies being carried out where those things are being assessed. So, there already have been early papers on detection of genome-wide methylation through re-sequencing—bisulphite re-sequencing of genomes."

As that technology matures, the detection of methylation and DNA-protein binding is also doable. But protein interaction is more difficult, as there isn't a way to comprehensively screen protein interactions in a high-throughput fashion for thousands of individuals in multiple tissues. Screening protein interactions is too artificial, he says, so researchers can't assess how those interactions are correlated throughout the whole system. Schadt would like to be able to assess when a protein is interacting with another protein and what other interactions are occurring simultaneously, but that's not possible with current technology. "You can only assess one interaction at a time and that's of limited use," he says.

SYSTEMATIC PERTURBATION

Schadt says of reconstructing networks based exclusively on gene expression data (his first attempt was a 2005 *Nature Genetics* paper), "you are almost doomed to fail as far as your ability to make predictive models because there are so many problems." Correlation-based information does not enable researchers to infer how things are causally related.

"In most cases, the different sub graphs you're considering are what are called 'Markov equivalents.' So, no matter how much data you had in that setting, you're not going to resolve the right structure because statistically they're indistinguishable." The key, he says, is to introduce a source of systematic perturbation to break that Markov equivalent. Groups have tried doing that gene knockouts or chemical perturbations, but Schadt says that's still artificial and one-dimensional.

By contrast, "The DNA variation data provide a naturally occurring source of systematic perturbation. Ultimately, it is DNA changes that drive how the system varies between species, between individuals within the species. So, it is the ultimate source of causal perturbation, but it's multi-factorial and occurring in a context where the complex phenotypes actually manifest themselves."

"This is the revolution of what we're doing—the DNA variation information has not systematically been leveraged as a source of perturbation to put together the causal network... It's only by including that kind of information that you can get to something predictive."

When Schadt's team tried constructing a probabilistic causal network on those yeast data using just gene expression data, it got "something that was completely not predictive. It could not predict the key causal regulators that we describe in [the *Nature Genetics*] paper. But when we layered in the DNA variation information along with protein DNA binding, we were able to make accurate predictions that we then tested prospectively."

BILLION DOLLAR QUESTION

The billion dollar question is whether Schadt's efforts will have a meaningful impact on Merck's drug discovery and development programs? "Obviously, we are motivated to apply all of these technologies toward that end," says Schadt. For example, roughly one-third of all the novel targets being pursued by the diabetes and obesity franchise—programs with medicinal chemistry support that are actually developing a drug—have stemmed from his group's efforts. "These are all completely novel, not characterized in the literature, just completely unexpected. You never would have locked onto them without doing the type of integrative thing that we did." And that's exactly the goal, of course.

"There's such a massive amount of data being generated, dumped into the public domain. The real question of the 21st century is, who are going to be the right kinds of groups to integrate that information in ways that best informs the

ANDY REYNOLDS

Schadt gets his data not just from DNA, but protein interaction and DNA protein binding.

biology at the earliest stages? [It] is an information-based game and who's going to emerge as the information brokers, the Google of the biomedical and life sciences, to tie that together?"

Whether people think the time is right today or five or ten years from now, Schadt says, "I think you're seeing a growing acceptance that that has to happen. There's no other way to mine the data."

Many people ask Schadt why his group has been so productive of late. "I get asked the question all the time. 'Why? You seem kind of academic. You're publishing these amazing papers. Why are you at Merck? Why aren't you at the Broad or something like that?'" Schadt responds that, given the scale of his group's studies, often involving thousands of patients (or mice) and multiple tissues, this work couldn't really have happened anywhere else. "It was really the support of Merck. Having a more visionary company and leadership like Peter Kim and Stephen Friend say, 'Man, we have got to better understand biology and we've got to understand it earlier in the process of drug discovery to increase our probability of success downstream.'"

MAN AND MACHINE

Aside from managerial support, there is a long list of credits. The computing architecture is a key aspect of this endeavor. "That's all part of the equation," says Schadt. "There's the money and the will to generate that scale of data, but then there's processing that data. What it takes to process, to build these kinds of models is non-trivial on the computational side."

Merck supported the building of a world-class, high-performance compute cluster—a 7000-CPU IBM Blade system—that ranks among the top 200 supercomputers. "It wasn't just cobbled together," says Schadt. "It was a very intentional design. It's a huge disk storage system integrated with [all] the software that's built around that to handle queuing and load balancing."

"The problems we're pursuing are as computationally demanding as anything you'll see in physics or climatology or any of the other fields... When we're reconstructing these probabilistic causal networks, there are 10,000 genes interacting with one another. That is the n-body problem."

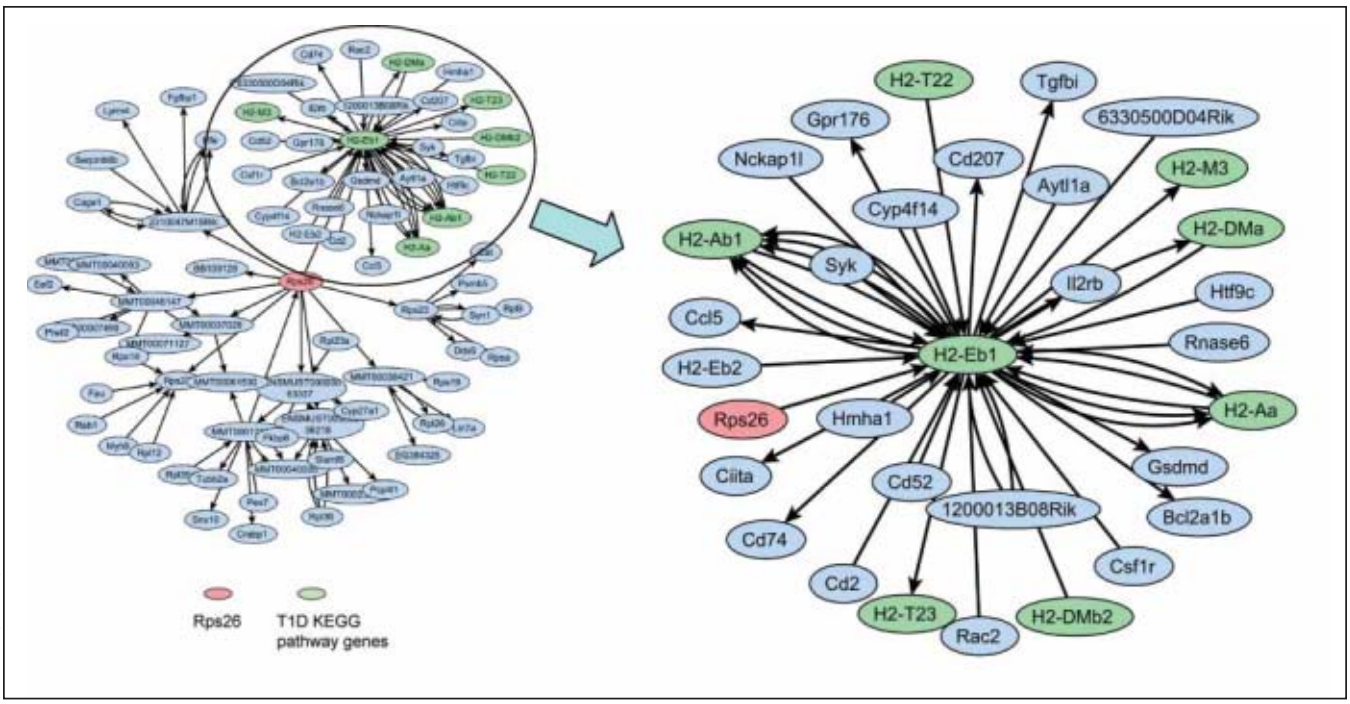
Besides the raw computing resources, Schadt pays tribute to a diverse group of about 40 researchers "who buy into a common vision that it is through these integrative methods and generating this scale of data and integrating it to produce these models that we can mine."

There's a network reconstruction group consisting of "mathematics, physics, computing type guys who are just driven to integrate data that way." The biology data mining group consists of "hard-core biologists working hand in hand with the network group to help put the networks together and then mine them." Other groups handle classic statistical genetics, data processing to feed the network reconstruction group, and bioinformaticians doing data integration. "The key, though, is that they're all working as one coherent team." Schadt also gives credit to Merck's "world class" informatics infrastructure that brings in the data and the wet lab biology groups that are close partners.

Schadt's group builds most of its tools either internally or in collaboration with other academic centers. But commercial efforts from companies like Gene Network Sciences, Genstruct, and Entelos are high on his radar. "We don't have any illusion that we're going to solve all the problems," he says. That's one reason his group is involved in the DREAM Project, for example.

Schadt says that Merck—largely through Kim—is encouraging an open approach to publishing "the methods and the thinking," because external validation complements what Merck can do internally. With the field moving so fast, it's crucial to receive advice from colleagues. "The strategy has always been to be very open with the kinds of methodological developments, how it gets applied and ... not think we're going to solve everything."

He particularly values some "great



PLOS BIOLOGY, MAY 2008, VOL. 6

Causal network reconstruction by Schadt's group showed *RPS26* was directly connected to known Type 1 diabetes pathway genes. A similar exercise for *ERBB3* shows no such enrichment.

systems type people" in academia such as Andrea Califano (Columbia), Daphne Kohler (Stanford) and Trey Ideker (UCSD). Obviously the openness has to stop somewhere. Says Schadt: "Those things that Merck decides to place its bets on we're not going to be out talking about, because that's the value Merck's going to realize. But everything that helped inform that decision you want to be open."

CROSS TISSUES

One of the advances Schadt is most excited about is to "model the entire system at the molecular level to get at networks that actually predict physiological states." Most of the current work focuses on a single tissue. But Schadt's group is starting to look at multiple tissues. "For example, in mouse, we can look at six to nine tissues from the same animal over hundreds of animals and start figuring out how these different tissues are communicating with one another via the molecular network."

For example, "how do changes in the hypothalamus induced by DNA variation change molecular states in the brain that go on to cause molecular state changes in adipose tissue or pancreas or liver or stomach?" Work on such cross-tissue networks will be one of the more exciting papers coming next: how behaviors in one tissue affect behaviors in another to drive toward a disease state.

For now, these models are mainly descriptive, but Schadt is driving toward applying the same techniques for single tissues with multiple band tissues. "We do have these concepts emerging of what we call module-to-module causality, where we're actually fitting models. So, a module would be just a sub network within a tissue and when you perturb that module with say a DNA change and that module changes, can we predict what other modules in other tissues are going to change?" A recent paper in collaboration with a group at the University of Wisconsin (*Genome Res.* 2008 May;18(5)) began to explore that idea—a network diagram "showing module-to-module or sub network-to-sub network connectivity both within and between tissues."

The Rosetta team hopes to be able to modulate that system to see what changes. Says Schadt: "We're not quite there yet, but I think that over the next couple of years that's exactly where we want to be." •

FURTHER READING:

Zhu, J. *et al.* "Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks." *Nat Gen* 2008

Schadt, E.E. *et al.* "Mapping the genetic architecture of gene expression in human liver." *PLoS Biol* 2008

Chen, Y. *et al.* "Variations in DNA elucidate molecular networks that cause disease." *Nature* 2008

Emilsson, V. *et al.* "Genetics of gene expression and its effect on disease." *Nature* 2008

Schadt, E.E. *et al.* "An integrative genomics approach to infer causal associations between gene expression and disease." *Nat Gen* 2005

Schadt, E.E. *et al.* "Genetics of gene expression surveyed in maize, mouse and man." *Nature* 2003

Timing and Other Data Storage Trends

Vendors tackle the problems of where to store, when to access, and how to archive data.

BY MIKE MAY

As scientists grapple with the deluge of data in virtually every facet of drug R&D, data storage companies are themselves racing to offer the best and greenest solutions (see, "The DNA Data Deluge," *Bio•IT World*, April 2008). Beyond the difficulty of simply storing that data somewhere, scientists need ways to work with it. A spate of offerings from leading data storage vendors addresses those issues and reveals that many of today's decisions depend largely on timing.

The best place to store data depends on when it will be needed as well as its age. It usually makes sense to put fresh data on fast-access storage, like a disk, and older data relegated to tape. For example, Deepak Thakkar, director of industry marketing at Silicon Graphics Inc. (SGI) in Sunnyvale, California, says that with SGI Data Migration Facility software, "we can orchestrate incoming data into different storage channels." With this software, a user defines rules that automatically organize stored data.

Tape storage is a must in some circumstances. Bruce Hillsberg, IBMs' director of storage system research in San Jose, California, says, "If you are storing billions of files and petabytes of data—and companies do that—then you need an effective and green way of doing that. Tape is a great choice." SGI also supports the greening trend in today's data storage. For example, when SGI's data migration facility puts information on tape, that storage requires zero power, unless the data are being accessed. So while today's scientists collect more information, they can (sometimes) do it with less energy.

Hot Hardware

Many of the biggest improvements in data-storage capabilities come from hardware-software combinations. In August, EMC (Hopkinton, Massachusetts) introduced its CLARiiON CX4 series, which feeds off several advanced technologies. For one thing, the CX4 series uses flash drives, which provide nonspinning, solid-state storage. According to Ruya Atac-Barrett, EMC's director of CLARiiON marketing, "These drives provide 30 times more processing power than spinning drives and response times are 10 times faster." She adds, "Flash drives are 98% energy efficient." One CLARiiON system can also store nearly a petabyte of data. Moreover, RSA enVision security software comes built-in with CLARiiON CX4 systems. This software package is produced by EMC's security division, RSA, and it provides a range of features, including reporting unusual usage and summaries of inbound and outbound traffic.

IMAGE COURTESY OF EMC.

EMC's CLARiiON CX4 series relies on a variety of high-tech advances, including flash drives.

Also in August, EMC released Celerra NX4, which is aimed at medium-size health care organizations. It can be used in a variety of configurations: network attached storage (NAS), Internet smart computer system interface (iSCSI), and fiber-channel storage area networks (SAN). The NX4 also provides 60 terabytes of storage.

In many life science research applications, changing technology requires data storage that can grow. Consequently, BlueArc (San Jose, California) adds Dynamic Write Balancing to its storage solutions. "When you add storage," says James Reaney, BlueArc's director of research markets, "this automatically rebalances the data across old and new storage. Otherwise, you would have to copy all of that data, reformat it, and then copy it back into storage."

Isilon Systems in Seattle, Washington, also kept growth in mind with its IQ X Series. "You can start with a simple system, and then add nodes as you need them," says Jay Wampold, Isilon's senior director of marketing and communications. "When you add a node to the cluster, the system automatically balances the data." In fact, Isilon says that its storage systems are so easy to use that it takes less work to manage data. Chris Blessington, Isilon's director of marketing communications, says that one customer "had been using six

full-time IT people just to manage storage. After switching to an Isilon system, they needed zero full-time employees to manage storage." (see, "Isilon Insights," *Bio•IT World*, April 2008).

In September, Isilon introduced an even more-advanced storage system, its OneFS (one file system) 5.0. "This delivers 20 gigabytes of aggregate throughput," says Wampold, "and it can scale up to 2.3 petabytes of storage in a single-file system."

Archiving for the Ages

There are certain cases where biotechnology and pharmaceutical companies need to keep some data for years, possibly even decades. Much of that data might include repeats, such as reports that use some of the same data. In addition, those data must be kept safe, even though they won't be used often. IBM provides a few new approaches to those problems.

For replicated data in stored files, researchers can turn to deduplication, which IBM's Hillsberg describes as "the ability to take data that you will store, look for repeating patterns, and then factor out those patterns." To move into deduplication faster, IBM acquired Diligent Technologies in Framingham, Massachusetts. "Diligent has the leading deduplication product, and it can reduce the space that information needs by a factor of 25, depending on the data," says Hillsberg.

> Imagine data stored in some virtual environment where many people could access the data simultaneously... and collaborate in real time. If we collaborate, that's good for all of us.

IMAGE COURTESY OF BLUEARC

Modules add storage to BlueArc's Titan, which automatically rebalances the data.

In July, IBM introduced its System Storage TS1130, which holds up to a terabyte of data. And for a tape drive, this one is fast—providing an input-output bandwidth of 160 megabytes per second. The tapes used for this system are smaller than two packs of playing cards. Moreover, data put on one of these tapes should last a decade, maybe longer. "It depends how you care for the tape," says Hillsberg. In addition, the TS1130 even includes data encryption. "If you lose a tape," Hillsberg says, "there's no way anybody could get data off it."

Despite the speed of the TS1130, it's not like disk storage in terms of access. "For data that you need in subseconds, store it on a disk," Hillsberg says. But how do you decide which data to put where? If desired, IBM's DR550 storage system will decide for you. "It comes with embedded disks, and it can use disk or tape storage that you already have in your storage network," Hillsberg says. Likewise, BlueArc's Data Migrator runs on a set of user-defined rules to decide where to store data. "This is very useful to get the data to where all of the researchers are," says Reaney.

Interacting with Data

"We tend to think of storage as just storage," says Thakkar, but he sees the future of storage as a collaborative entity. Imagine data stored in some virtual environment where many people could access the data simultaneously. "People could collaborate with each other's data in real time, comment on it, and then merge together different data sets," he says. "If everyone just works in their own silos then drug-research programs are only as good as one person's thought processes. If we could collaborate, that's good for all of us."

Another task ahead involves so-called logical preservation, which means being able to open a file that is decades old and still having software to read and manipulate the data. "IBM is working on this in our research division," says Hillsberg. "It requires figuring out the best way to do this without losing data because of changes in applications and formats."

So time plays a fundamental role in a range of data-storage characteristics. Time impacts the best format for storage, and as time passes, the data must remain safe and readable. Yet, staying on top of the rushing ticks of data collected keeps storage scientists very busy. •

Bottlenecks in Drug Discovery

Breaking down silos is a key business initiative.

BY FRANK KLEIN

Amid a more stringent regulatory environment and soaring R&D costs, the imperative is to "fail fast and fail cheap" with new drug candidates. While there is no doubt that the pharmaceutical industry is growing in complexity, there are relatively simple steps companies can make to address logistical bottlenecks throughout the drug development cycle.

Throughout the development of a candidate drug, research teams work in isolation on tightly-defined projects, often operating independently, or even competitively. This lack of cooperation is exacerbated by the increasing complexity of tests and treatments available and a frequent lack of formal structures for sharing intelligence. This results in the development of information "silos" in which researchers are not able to access all of the intelligence previously gathered in a candidate drug's development, and ultimately, the repetition of expensive research.

"A siloed approach to the adoption of technology has left the majority of pharmaceutical organizations with some serious business and technological challenges ahead," wrote Markella Kordoyanni, a pharmaceutical technology analyst at Datamonitor, in the 2007 report, Trends to Watch: Pharmaceutical Technology. The objective should be to improve communication channels between laboratory scientists and those working in the preclinical and clinical stages of drug development, so that scientists can use clinical feedback to refine their investigations. For example, reliable comparisons of *in vivo* and *in vitro* safety study data linking subjects' symptoms with molecular data creates a bridge between classical and molecular pathology. Biomarkers identified in the laboratory can be tracked into clinical trials—or vice-versa—providing clear proof of mechanism and proof of concept. As translational and personalized medicine become a reality, open communication between all stakeholders is critical. One area in particular in which information processing and sharing is becoming increasingly critical is proliferating digital imaging data.

Image Backup

Today, more than 70 per cent of all data generated in life sciences research is in the visual form, but its huge potential is only partially tapped. Pharmaceutical companies developing new drugs and therapies need to assimilate not only molecular and genomic information, but also complex, heterogeneous image data from all stages of the drug development cycle. In 2004, at the Fifth National Forum on Biomedical Imaging in Oncology, Janet Woodcock, director of FDA's Center for Drug Evaluation and Research commented that, "There is tremendous potential for the use of imaging in drug development... from pre-clinical [applications] all the way to using surrogate markers for approval." The drug development process is "increasingly challenging, inefficient, and costly," and imaging technology remains "at the forefront of [the FDA's] efforts to streamline the drug approval process."

Manual image analysis is labor-intensive and compounded by a shortage of experts to interpret digital data. It is not unusual for major pharmaceutical companies to have over 100 different imaging instruments, and their image analysis requirements are likely to increase exponentially in the coming years. Cell proliferation and immunohistochemistry studies are a critical focus of pre-clinical pathologists. These studies can prove time-consuming, requiring scientists to manually select regions of interest and conduct cell counts. Utilizing automated image analysis technology in a 13-week oral toxicity study on 180 mice, a major pharmaceutical company reduced the time spent on the analysis of a total of 9,000 images from 16 weeks to just four. Such inefficiencies are compounded through each stage of the drug development process.

While no system can supplant highly trained specialists, adopting systems that can help bring objectivity and automation to image analysis can help streamline research and the development of more effective therapies. Currently, in most bioscience and pharmaceutical companies, image analysis tools are purchased as point solutions to solve specific problems, often imbedded in the image acquisition tools themselves. Although they may improve the speed and accuracy of analysis, many accept only particular inputs or file types, restricting their usefulness. Replacing point solutions with a single, standardized platform that handles all of an organization's image analysis tasks can help alleviate productivity loss from inefficiencies in the analysis of digital data and expedite early candidate attrition rates.

Adopting systems that bring objectivity and automation to image analysis can help streamline research and the development of more effective therapies.

The adoption of integrated software infrastructure for the storage and sharing of information between researchers and clinicians can be a costly and intensive process. Nonetheless, the initial cost and time outlay are justified. Implementing appropriate software platforms removes a key hurdle in the occasional breakdowns in communication that exacerbate productivity loss and waste research funds.

Frank Klein is VP of medical imaging at Definiens. He can be reached at fklein@definiens.com.

Clinical Research

Pacific Rim Warms Up to EDC

With evolving language capabilities, EDC is now open to nearly everyone.

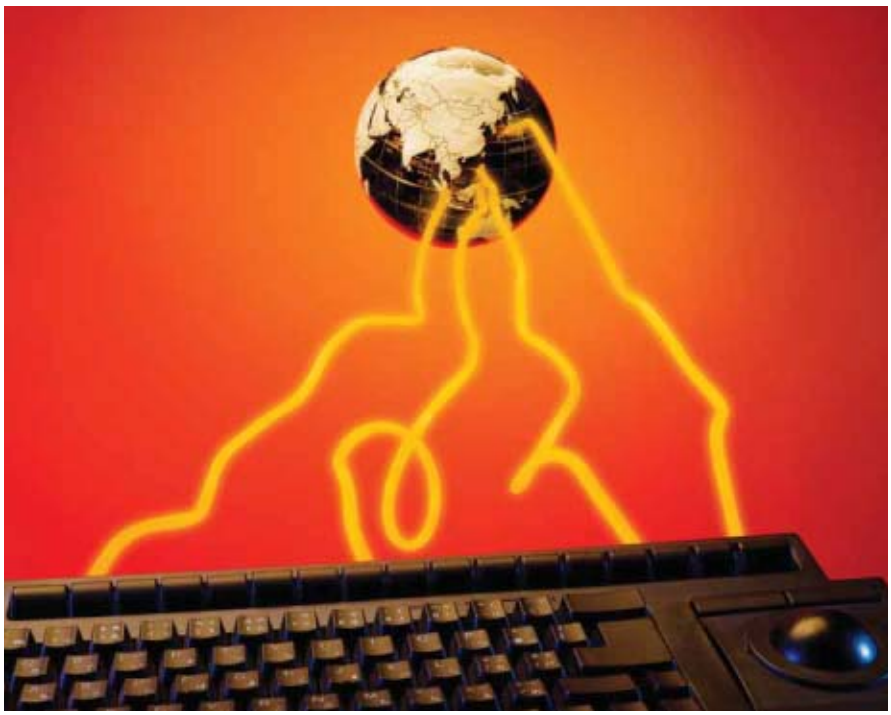
BY ANN NEUER

The use of electronic data capture (EDC) in the Asia-Pacific region is poised to equal or surpass more developed regions of the world, says Graham Bunn, vice president of global CRO partnerships at Medidata Solutions. "The exciting thing about the Asia-Pacific area is that it has traditionally been a lot further behind in EDC adoption but it seems to be catching up with Europe quite rapidly because sponsors are starting to understand and appreciate its advantages," he says.

Specifically, Bunn is referring to the fact that EDC has evolved to the point that users around the globe can enter data in their own languages, and those data are combined into a single database, adding to project efficiency. This is exceedingly difficult with paper trials because paper case report forms in multiple languages have to be translated and typically cannot be handled in one database. Being able to enter data in various Asian languages brings an important level of accuracy to the trial and comfort to new users who are experiencing EDC for the first time. "In the Asia-Pac region, clinical trials generally cannot be run in English, so it is important to enable data collection in the native language," Bunn says.

Another key advantage is the immediacy of EDC. With multiple time zones across Asia, sponsors and contract research organizations (CROs) can instantly see the progress of clinical trials, identify sites that are struggling with subject recruitment, and more easily include Asia-based staff and investigators in study-related communications and conference calls. Sponsors and CROs also appreciate the potential of EDC to reduce the number of costly monitoring visits—not a small thing in this vast multicultural region.

The move to EDC brings challenges, and according to Bunn, perhaps none bigger than the variable infrastructure across the Pacific Rim. "Infrastructure is definitely a challenge. In Singapore, 98 percent of everywhere has PC and fast Internet connections, and in Japan, the Ministry of Health, Labour and Welfare has launched a five-year national plan to expand use of technology in clinical trials based on CDISC standards. But in other parts of Asia-Pac, it is comparable to Latin America and how it was in Eastern Europe two or three years ago. Getting good Internet access and having efficient PCs can be a problem," he explains.



MARK GABRENYA

Culture Challenge

Training can also pose difficulties. Many Asian sites are not familiar with EDC, so sponsors and CROs are turning to state-of-the-art EDC systems with built-in multi-language eLearning capability that does not require high bandwidth. Potential users can be restricted from accessing the system until they pass online tests about the technology. A number of marketed EDC systems offer this function.

And finally, there is the issue of culture. "This is one of the big ones," says Bunn. Countries have different approaches, different languages and etiquette in running clinical trials, and this impacts how various Asian countries may view and accept clinical trial technology. The situation in Japan is changing quite dramatically as the International Conference on Harmonisation (ICH) lays the foundation for studies in the U.S., the E.U., and Japan to operate under harmonized guidelines and standards. Other countries in the region, however, are not yet aligned with Good Clinical Practice (GCP) guidelines. Bunn says, "Investigators have less clinical trial experience in some of these countries. In fact, some investigators are clinical-trial naïve, so they need training in running a clinical trial, handling the product, and EDC. They also need standard operating procedures. Locally based CROs can be enormously helpful in this effort."

Clinical Research

Refining Patient Safety

SAS takes a lifecycle view of patient safety data.

Jason Burke hopes the SAS solution will eliminate the need to manually implement safety algorithms.

BY ANN NEUER

The pharmaceutical industry and regulatory agencies are sharpening their focus on patient safety through intense efforts to improve the way safety data are collected, analyzed, and interpreted. And there is a long way to go, says Jason Burke, global director of the health and life sciences market segments at SAS. "Today, most organizations do separate analyses for drugs before and after approval. This is a challenge because it may not be obvious how the lifecycle of a drug looks when safety analyses are conducted using separate pre- and post-approval processes and systems," Burke explains.

A comprehensive lifecycle view of the safety performance of a drug is what is needed to provide a more repeatable, standardized, and auditable approach for exploring safety signals. The FDA is steering the market in this direction with the release of various guidances intended to improve how the agency disseminates safety information, how adverse events are to be reported, and how reviewers are to conduct the clinical safety reviews for the new drug application (NDA) and biologics license application (BLA) review processes. There is also the Sentinel Initiative, launched by the FDA in May to develop a national strategy for monitoring medical product safety, a nod to the emerging science of safety.

"The agency is driving toward getting a better understanding of the comprehensive view of what safety looks like over the entire history of the drug," Burke says.

With all this tailwind, SAS is on board with its just-released SAS for Patient Safety, a comprehensive solution designed to help users comply with recent FDA guidances by offering advanced analytics for signal detection and pharmacovigilance. The solution consists of a collection of SAS software that is implemented in conjunction with consulting services. It offers capabilities such as standardized safety reporting that leverages standards from the Clinical Data Interchange Standards Consortium (CDISC) and FDA guidances, visualization capabilities that enable researchers to understand patient safety data, and automated signal detection of published as well as SAS-developed signal detection algorithms. "The intent is to remove the need to manually implement commonly used safety algorithms by rolling them into the solution," Burke comments.

Responding to Change

An important feature of SAS for Patient Safety is that it allows for data aggregation and data integration, to enable standardization and to bring together information from disparate sources into a consistent repository. According to Burke, pharmaceutical companies typically rely on transactional safety systems within silos to produce periodic reports. "This siloed approach creates challenges for companies in terms of aggregating information from many sources, cleaning the data, and using standardized structures to report them," he says.

The technology behind the integration capability is the newly launched SAS Clinical Data Integration server, a platform that defines and automates processes for aggregating clinical data through the use of standards such as those of CDISC. "With this platform, SAS can provide a bridge across whatever sources of safety information a pharmaceutical company might be using to create a 360-degree view of the profile of the drug. We created SAS for Patient Safety in response to what we have seen as a sea change in the industry focusing on safety," Burke says. •

New Products

Transcriptome Translation

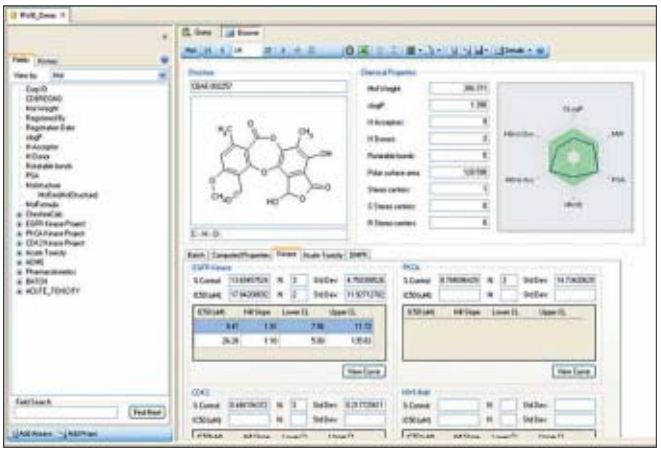
Applied Biosystems SOLiD Whole Transcriptome Expression Kit is a sequencing-based molecular tool for the genomic analysis of whole transcriptomes. The kit provides new views of a cell's transcriptome, such as expression of non-coding RNAs, identification of alternative splicing events, expressed SNPs or mutation, or translocations and fusion transcripts. The kit is based on Ambion technology.

Company: Applied Biosystems
Product: SOLiD Whole Transcriptome Expression Kit
Available: Now
For more information: solid.appliedbiosystems.com

Proteomics Software

Thermo Fisher Scientific and Sage-N Research have released the first proteomics software platform designed specifically for the Enterprise Linux market. SORCERER Enterprise is a scalable software suite for fully automated, high-volume proteomic analyses on high-performance Linux systems, including blade servers and conventional Linux clusters. It's designed for labs that need to process increasing amounts of proteomics data from mass spec experiments. The software can support hundreds of proteomics researchers using a web interface.

Company: Sage-N Research
Product: SORCERER Enterprise
Available: Now
For more information: www.sagenresearch.com



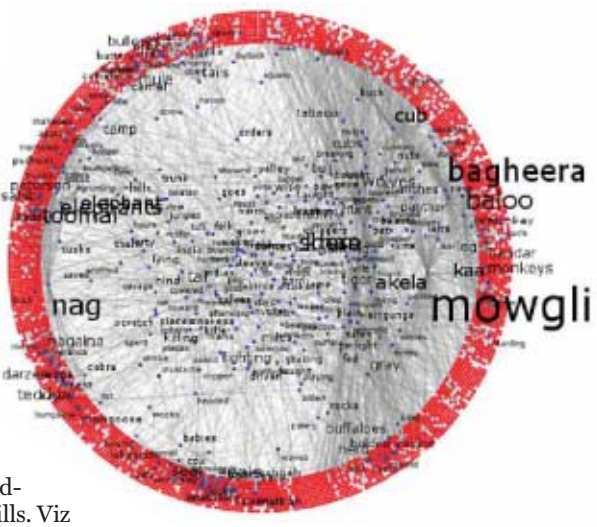
Informatics Update

Symyx has released Isentris 3.1, an informatics system with upgrades that include the ability to create, manage, and share fully searchable local databases with insert, update, and delete capabilities. Highly interactive views allow scientists to filter, sort, and cluster data themselves. 3.1 also offers support for local databases, molecule and reaction clustering, chemical structure searching in Excel spreadsheets, molecule- and reaction-based transfer of results into Excel, and writable .NET controls for inserting, updating, and deleting data.

Company: Symyx
Product: Isentris 3.1
Available: Now
For more information: www.symyx.com/isentris31

Spontaneous Data Viz

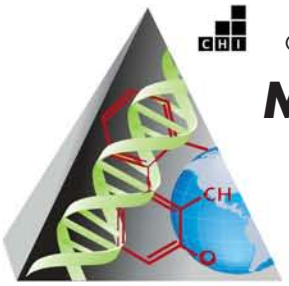
SPSS Viz Designer creates dynamic visualizations, and can intelligently recommend the most appropriate chart or graph so the data can be communicated in the most understandable way. With an intuitive, graphical, drag-and-drop design interface, Viz Designer eliminates the need for advanced programming skills. Viz Designer uses nViZn, a patented SPSS technology based on "The Grammar of Graphics" principles from Leland Wilkinson, and one of the most sophisticated visualization technologies available.



Company: SPSS
Product: Viz Designer
Available: Now
For more information: www.spss.com/VizDesigner

Advertiser Index

This index is provided as an additional service. The publisher does not assume any liability for errors or omissions.



Cambridge Healthtech Institute's 16th International

Molecular Medicine Tri-Conference

Shaping Future Medicine

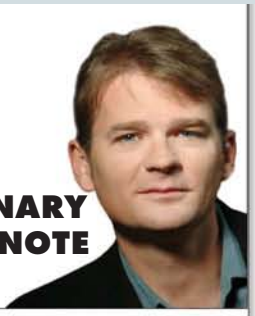
Conference Dates: February 24-27, 2009
Moscone North Convention Center

Exhibit Dates: February 25-26, 2009
San Francisco, California

UNLOCK THE SECRETS OF YOUR DNA!
30 People That Register by October 24, 2008 will Receive a FREE 23andMe Custom Genome Scan. See Tri-Conference.com for details

23andMe *(a $395 value)*

- **Molecular Diagnostics**
- **Mastering Medicinal Chemistry**
- **Stem Cells Congress**
- **Executive Summit on Strategy and Innovation**
  Qualified Attendance
- **Preclinical Drug Safety**
- **Preclinical Development of Biologics**
- **Translational Medicine**
- **Cancer Profiling & Pathways**
- **Adopting Integrated R&D Informatics Systems**
- **Cancer Molecular Markers**
- **Companion Diagnostics**



PLENARY KEYNOTE

Brave New Age of Personalized Medicine

David Ewing Duncan, Chief Correspondent of NPR Talk's "Biotech Nation" and best-selling Author, "Masterminds"

Corporate Sponsors

CHEMICAL COMPUTING GROUP INC.
INGENUITY SYSTEMS
nanoString TECHNOLOGIES

Lead Sponsoring Publications

Bio-IT World
GEN Genetic Engineering & Biotechnology News
nature
Science AAAS
TheScientist

Co - Sponsors

The State of California
THE WALL STREET JOURNAL

CHI
Cambridge Healthtech Institute
250 First Avenue, Suite 300
Needham, MA 02494, USA
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 U02 when registering!

Tri-Conference.com

Educational Opportunities

This section provides a variety of educational events in the life sciences industry that will help you with your business and professional needs. To list an educational event (print & online), contact **Sales_chmg@chimediagroup.com**. To preview a more in-depth listing of educational offerings, please visit **www.bio-itworld.com**

Featured Events

Harvard Biomed HPC Leadership Summit, Oct 5–7, 2008, Boston, MA
This summit is the premier event for professionals who build and maintain clusters and grids for high performance computing in the biomedical fields. Contact *biomedhpcadmin@hms.harvard.edu*

Drug Stability Reporting (eStability) Working Group Vendor Meeting, Oct 7, 2008, Rockville, MD
This meeting will inform vendors (LIMS and others) of the Drug Stability Reporting (eStability) standard for the transfer of stability data electronically and to generate interest in developing methods of transforming stability data into a format compliant with the proposed HL7 XML schema. Contact *Norman Gregory at norman.gregory@fda.hhs.gov*

DIA 2nd Annual Clinical Forum: Data Driven Drug Development Decisions, Oct 20–23, 2008, Ljubljana, Slovenia
A gathering of over 700 of the world's premier clinical research professionals from over 20 countries, the DIA Clinical Forum will again provide broad learning, discussion and networking opportunities. For more information or to register, contact the DIA Europe Customer Service Team at *+41 61 225 5151, M–F, 8:00–17:00 CET*, or email *diaeurope@diaeurope.org*

15th Annual Biopharmaceutical Applied Statistics Symposium (BASS), Nov 3–7, 2008, Savannah, GA
BASS offers tutorials and short courses on diverse topics pertaining to the research, clinical development, and regulation of pharmaceuticals with speakers from academia, the pharmaceutical industry, and the FDA. Don Berry will provide the keynote address on "New Developments in Bayesian Clinical Trials." For further information, visit *http://bass.georgiasouthern.edu* or contact Ruth Whitworth at *912-478-7904 (rewhitworth@georgiasouthern.edu)* or Laura Gunn at *912-478-7422 (lgunn@georgiasouthern.com)*

The Next Tool for Healthcare Innovation, Nov 20, 2008, 2pm EST
A complimentary Frost & Sullivan eBroadcast. In an ever shrinking world, webcasting has developed into a very effective tool in the healthcare space as it allows healthcare professionals to overcome the difficulties that come with travel and distance. Efficiencies in cost reduction using webcasting have been well demonstrated and documented. Improvements in technology can only help to cement these applications in the healthcare industry, including physician/employee education and clinical trial tracking. Visit *http://ebroadcast.frost.com/hcinnovation*

CHI Events

For more information on these conferences and other CHI events planned, please visit **www.healthtech.com**

Adaptive Clinical Trials
October 23–24, 2008 • Washington, DC
Bridging Pharma & IT
October 28–30, 2008 • Providence, RI
Bridging the Business Development/ Alliance Management Interface
November 5–6, 2008 • Boston, MA
Post-Approval Drug Safety Strategies
November 12–14, 2008 • Arlington, VA
Strategic Resource Management
November 17–18, 2008 • Philadelphia, PA

Barnett Educational Services

For more information on Barnett live seminars and interactive web seminars, please visit **www.barnettinternational.com**

Live Barnett Seminars

Working with CROs
October 23–24, 2008 • San Diego, CA
Patient Recruitment and Retention
November 18–19, 2008 • Philadelphia, PA
Pharmacokinetics: A comprehensive overview of principles and applications
November 20–21, 2008 • Philadelphia, PA

Interactive Web Seminars

Electronic Medical Records & Source Document Verification
October 27, 2008
Innovative Preclinical & Clinical Strategies to Accelerate & Optimize Phase I–II Clinical Trials
October 31, 2008
New FDA 2008 Guidance: How to Complete the FDA Form 1572
November 10, 2008
The Pharmaceutical & Medical Device Industries Today
November 20, 2008

White Papers, Webcasts, and Podcasts

Keep abreast of current industry trends and developments. Browse our extensive list of free Life Science white papers, webcasts, and podcasts on **www.bio-itworld.com**

White Papers

Managed Innovation, Assured Compliance: *Developing, executing, and managing the transformation, analysis, and submission of clinical research data with SAS Drug Development*
Sponsored by SAS
White paper covers how to: Assemble data to foster better collaboration; Obtain up-to-date information during clinical trials; and Make informed decisions earlier in the trial process.

Addressing Life Sciences' Constantly Growing Data Challenges Research Environments
Sponsored by BlueArc
White paper covers: Factors driving the data explosion in the life sciences; New data management issues that must be addressed; HPC trends that are placing new demands on storage; and Storage solution attributes that address performance, manageability, and energy efficiency.

Webcasts

Improve Patient Care and Eliminate Paper Process: *Will Healthcare IT Trends Standardize?*
Sponsored by Adobe
Webcast covers: The current industry trends and the challenges these trends create; How to use best practices to improve patient care, enhance productivity and eliminate paperwork; How to reduce processing costs, eliminate paper costs, shorten process cycle times, and provide better coordinated care; The successful implementations at Beth Israel Deaconess Hospital, St. Vincent's Hospital, and Lahey Clinic.

Storage for Science – Methods for Managing Large and Rapidly Growing Data Stores in Life Science Research Environments
Sponsored by Isilon
Webcast covers: Guidance on the many storage requirements common to Life Science research; The evolution of modern data storage architectures; Major data storage architectures currently in use; and Isilon's IQ clustered storage product as a strong and flexible solution to those needs.

Podcasts

Hospital Paperwork No Longer has to be an In-Patient Procedure
Sponsored by Adobe
Podcast covers how Adobe solutions for healthcare can help you streamline your paperwork and stop making paperwork an in-patient procedure.

The Russell Transcript

Cancer Studies Suggest Pathways are Best Targets

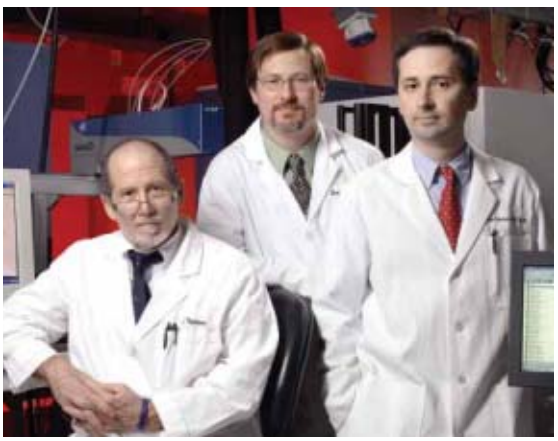
JOHN RUSSELL

Last month, two papers in *Science** described the integrated genomic analysis of pancreatic cancer and a form of brain cancer, Glioblastoma Multiforme (GBM). They not only revealed important biology but also suggested that the pharmaceutical and health care industries must fight solid tumors differently—principally by aiming at pathways rather than individual targets and focusing on early detection rather than late stage cures.

The studies, led by researchers from Johns Hopkins, examined roughly 20 tumors of each cancer type. They sequenced all protein-coding genes (~21K) and examined expression patterns and copy number changes. What they found was tremendous genetic diversity. A typical pancreatic cancer tumor had 63 genetic alterations; on average 49 of those changed the genes and their products.

Results were similar for GBM where a typical tumor contained about 60 genetic alterations. Interestingly, the GBM work identified two different sets of genetic changes and different mechanisms at work suggesting GBM is really two different diseases, not one as has been thought.

This picture of solid tumor heterogeneity poses a potential nightmare for selecting individual targets. However, when the genetic changes were put into pathways, a different picture emerged. In pancreatic cancer, investigators identified a core of 12 altered pathways that each individually affected over ⅔s of the tumors analyzed. Similar pathways were found to be altered in GBM as well as a few others.



From left to right, Vogelstein, Kinzler and Velculescu.

These results were considered so striking that a press teleconference was quickly arranged (9/3) the day before online publication in *Science* attended by paper authors and cancer researcher heavyweights: Bert Vogelstein, co-director of the Ludwig Center for Cancer Genetics and Therapeutics (LCCGT) and Howard Hughes Medical Institute; Victor Velculescu, associate professor of oncology, Johns Hopkins Kimmel Cancer Center, and Ken Kinzler, co-director of LCCGT and HHMI.

Cutting to the chase, Vogelstein said, "It is extremely unlikely that drugs which target a single gene, like Gleevec, will be active against a major fraction of solid tumors. Instead of screening for dugs against single proteins our work suggests it may be more productive to screen for drugs that act on the core pathways that are dis-regulated in most cancers."

"By targeting the pathways, it's possible new drugs could be effective against a much greater fraction of tumors. This is a very different perspective from what's now operative in the drug development community," said Vogelstein.

Secondly, said Vogelstein, "I think it's accurate to say that 99% of applied cancer research now goes toward developing new therapeutics. It is, I think, apparent from studies like ours that it is going to be even more difficult perhaps than previously expected to derive real cures from such therapies. One interpretation of our work is that the proportion of effort and funding devoted to other ways of managing cancer, such as prevention and early detection, should be greatly increased... as they may have much more success in minimizing cancer deaths."

It is important to note that these kinds of studies are made possible by tools that are only now starting to prove their value.

Pathway analysis tools, for example, were important in this work; indeed, GeneGo not only provided tools for the cancer work discussed here, but also conducted much of the analysis, and landed authors—Tatiana Nikolskaya (president & CSO) Yuri Nikolsky (CEO)—on both papers.

More tools are needed. Vogelstein, for example, suggests new imaging technologies to detect activated pathways must be developed. He further predicts that a simple blood test for some cancers is no longer science fiction

"We now know precisely how many genetic alterations in coding genes there are in typical pancreatic and brain tumors, for example, and with current technologies it's actually easy to detect many of them in the cancers and it will be possible soon to detect many of them on other samples from patients, say in their blood."

*Williams Parsons, D. *et al.* "An Integrated Genomic Analysis of Human Glioblastoma Multiforme." *Science*, published online 4 September 2008; Jones, S. *et al.* "Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses." *Science*, published online 4 September 2008

I WANT

TO WORK ANYWHERE I CHOOSE

First backlit keyboard for business.
Dell Connection Manager. Up to 19 hours of battery life.

THE NEW LATITUDE™ E6400

DELL.COM/IWantSimplicity 800.727.4247