Bio·IT World
Technologies Driving Discovery, Development, and Clinical Trials

Welcome to your digital edition of *Bio-IT W*

We hope y io-IT World community, er's Guides, Events, and OM for daily delivery of f ines on your desktop (or y

In addition, overing such targeted to Regulatory Compliance olete list of enewsletter gn-up.html

And as alwa l us an email with DIGITA

>> CLIC
>> CLIC

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

Oct 2007 Issue	
Total qualified digital subscribers:	7186
Biopharma/Chemicals	2176
Biotech/Bioinformatics	1834
Acad/Gov, Clinical Research	1854
Hosp Lab/Healthcare, Other	1322
Sept 2007 Issue	
Total emails sent:	7285
Undeliverable:	82
Total delivered:	7203
Removed:	21

NAVIGATION KEY

- Previous/Next Page: mouse click (or keyboard) on left/right arrows to turn page
- Contents: takes you to the Table 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 phase (can also search archives!)

I CAN:

LEAD WORLD-CLASS TEAMS TO FURTHER REVOLUTIONIZE DRUG DEVELOPMENT.

I CAN'T:

QUICKLY IDENTIFY, ATTRACT AND SCREEN HIGH-IMPACT CLINICAL TALENT WITH SPECIALIZED THERAPEUTIC EXPERIENCE.

*Survey of Yoh customers 2005-2006. © 2007 Yoh Services LLC – A Day & Zimmermann Company

WE CAN. Yoh has been delivering high-impact talent to America's leading companies for the past seven decades – longer than any of our competitors. These customers count on our experience, time-honored processes, responsiveness and market intelligence. And our delivery of talent that is 24.7% better than the competition*.

Focus on your job. And your future.

YohClinical
Know greater talent.®

yohclinical.com

Cambridge Healthtech Institute

www.bio-itworld.com

Bio·IT World

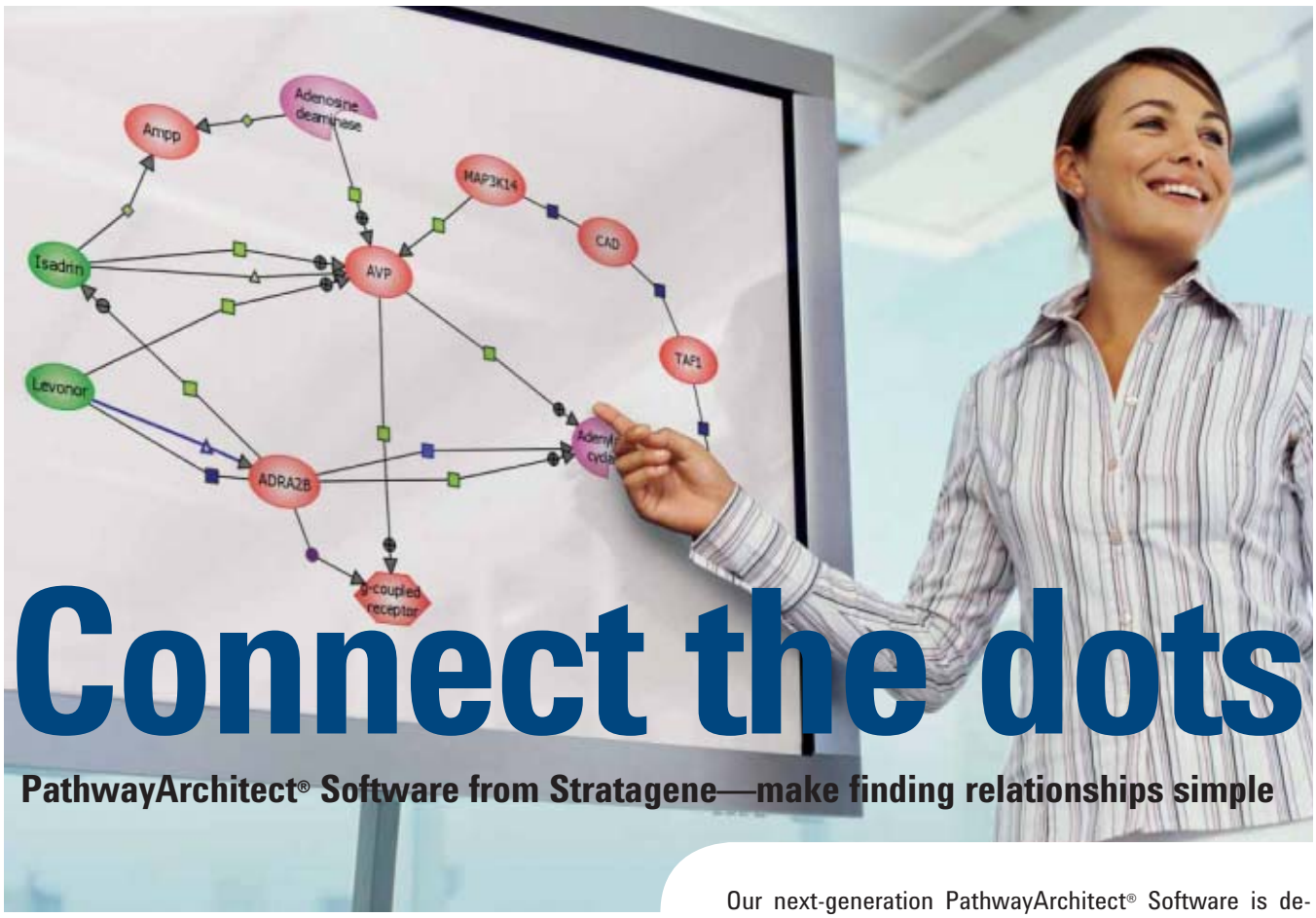
Indispensable Technologies Driving Discovery, Development, and Clinical Trials

OCTOBER 2007 • VOL. 6, NO. 8

DATA MANAGEMENT: THE NEXT GENERATION

Informatics and IT solutions for the exploding data crisis 26

Rico Magsipoc, CTO at UCLA's Laboratory of Neuro Imaging.



Connect the dots

PathwayArchitect® Software from Stratagene—make finding relationships simple

TRY IT FREE TODAY!

Download a trial version today at www.stratagene.com/softwaresolutions

Our next-generation PathwayArchitect® Software is designed for the life scientist who needs to obtain facts about a disease-relevant biological pathway as well as understand the statistically relevant relationships in their system and integrate complex numerical data. The guided workflow within the software takes you through the steps of searching for your pathway of interest. You can also build a pathway based on a list of genes and utilize the software to integrate and analyze your numerical data.

- **Pathway database expands target discovery**
- **Uncover new biological relationships**
- **Visualize your expression data as pathways**

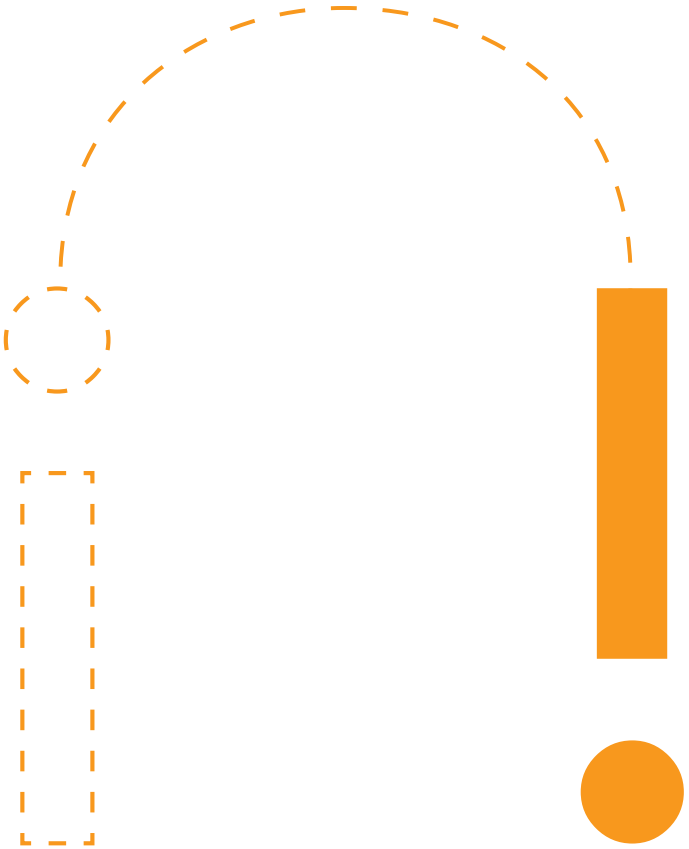
Visit www.stratagene.com/softwaresolutions for complete details about our PathwayArchitect® Software and information about Academic, multiple year, and multiple user licensing options—and connect the dots.

u.s. and canada 800-424-5444 x3
europe 00800-7000-7000

www.stratagene.com

STRATAGENE
An Agilent Technologies Company

Strand Life Sciences



Turn sequencing on its head.

It's here. Now you can perform a range of genomic experiments that you only dreamed of before. The Illumina Genome Analyzer changes the way you approach every kind of genomic experiment. The discoveries are happening, the wait is over, and our customers are publishing their results.

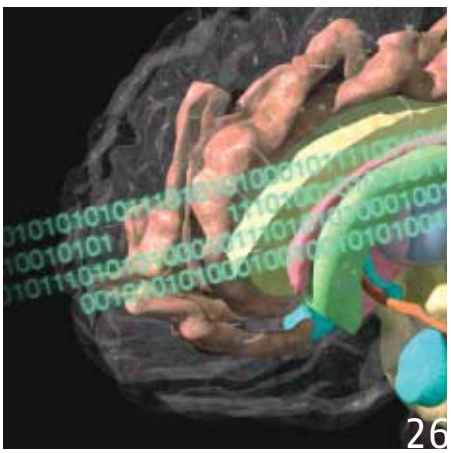
Change the way you think about sequencing.
Join the growing Illumina Community.

Do more with sequencing today:

www.morethansequencing.com

illumina®

Contents [10•07]



Cover Story

Data Management: The Next Generation

New technologies aim to simplify data management while meeting life sciences' soaring computation performance requirements. 26

Up Front

Computational Biology

Computational Development

IT/Workflow

Clinical Research

In Every Issue

Cover photo by Edward Carreon

Alternative Thinking About Server Rooms:

ADDITION THROUGH SUBTRACTION.

Take away the jungle of cables. Take away the so-called "normal" energy requirements and maintenance. What do you have? Introducing the HP BladeSystem c3000. All the technology of our larger BladeSystem in an efficient, compact, affordable package.

Technology for better business outcomes.

Powered by the Quad-Core Intel® Xeon® Processor[1]

See how less is more. Visit hp.com/go/nocompromise21
1-888-335-6443

1. Intel, the Intel logo, Xeon, and Xeon Inside are trademarks or registered trademarks of Intel Corporation in the U.S. and other countries.

First Base

From Best Practices to Navigenics

KEVIN DAVIES

In June 2005, a soft-spoken researcher from TGen (Translational Genome Research Institute) flew from Phoenix to Washington D.C. to accept a *Bio•IT World* Best Practices Award* for basic research. Dietrich Stephan was then head of the Neurogenomics division at TGen, searching for complex disease genes including autism and amyotrophic lateral sclerosis. The award recognized his team's success in deploying new software to map a gene for sudden infant death syndrome.

We rather lost track of Stephan after that, but were delighted to learn this summer that our humble Best Practices winner had co-founded an exciting new consumer genomics company, Navigenics. Whereas 23andMe (See "23andMe," *Bio•IT World*, June 2007), is laying low for now (co-founder Linda Avey will keynote *Bio•IT World* Expo in 2008), Stephan was only too happy to shed light on Navigenics' mission**, and that he still proudly keeps his Best Practices trophy on his desk.

Dietrich Stephan (left) receiving his Best Practices Award from Editor Kevin Davies in 2005.

"The Holy Grail is to sequence the whole human genome, put it in a big computer, push a button and have a rank ordered list of disease predispositions pop out the back," Stephan told me by phone. "The problem is we've never been enabled with the technology to resequence the genome in a cost-effective and accurate way. We still don't have that ability. But we can look at common human diseases caused by common genetic variants floating around the population at large."

Stephan began planning a company two years ago (perhaps buoyed by his *Bio•IT World* award) recognizing the rapidly improving capacity of SNP chips to conduct genomewide surveys. He was introduced to co-founder David Agus (director, Spielberg Family Center for Applied Proteomics, Los Angeles), who had similar ideas. "In a five-minute conversation on the phone, we had the same vision; we banded together," recalls Stephan.

They secured funding from two leading VC firms, and began assembling experts in ethics, legal affairs, genetic counseling, science, and epidemiology. There is also an editorial team tasked with communicating delicate and sophisticated information to physicians and a lay audience.

SNP Service

Stephan says Navigenics will debut early next year using the new Affymetrix 6.0 chip, which contains 1 million SNPs. Eventually, the company will offer whole-genome sequencing. "We really see this as, at some point in your lifespan, everyone will get their genome sequenced and get risk mitigation. [Think of it as] newborn screening for adults." As for potential sequencing partners, he says, "no one knows which technology will win... We'll wait for the technology to stabilize."

But Stephan is adamant that whole-genome sequencing will offer much more information than a catalogue of SNPs. He points to recent findings in prostate cancer and heart disease, where disease loci were tracked to regions bereft of genes. "If you just sequence annotated regions, you're not going to get all the [medical] information," he says. Take autism: "We currently think autism is caused by private mutations and we haven't found a whole-genome association signal to date. So autism might be teased out using a sequencing strategy rather than a genotyping strategy."

Stephan has tested himself, of course. He does not carry an Alzheimer's-associated ApoE4 allele, but does carry some breast cancer predisposing mutations, which coupled with his family history, "might [require me] to get a checkup." In all, Navigenics will be launching for 20 conditions, using "a very intuitive dashboard" to communicate health risks based on genotype and advice for managing health proactively. Says Stephan: "It will encompass lots of information, supported in specific ways, taking into account average lifetime risk, how you're loaded relative to the general population, age of onset, a number of different nuances."

After conversations with 23andMe co-founders Anne Wojcicki and Linda Avey, Stephan views the two new companies as very complementary: "Navigenics is focused squarely on medical risk assessment on actionable conditions. 23andMe is focused on ancestry genealogy and that's going to be their space." Probably not for long, however. Jay Flatley, CEO of 23andMe's genotyping partner, Illumina, told analysts last month, "I can say that [23andMe is] going to be focused ... on the medical aspects of genotyping," although the early emphasis will be more on genealogy.

---

* *Bio•IT World's* Best Practices Awards are back : see p. 35.

** For the full interview with Dietrich Stephan, go to www.bio-itworld.com.

Count on Us.

Over 1500 customers worldwide already count on Thermo Scientific informatics solutions to manage one of their most important assets – laboratory data. These industry leaders have turned to Thermo Scientific LIMS and CDS software in their efforts to make the world healthier, cleaner and safer.

If you are looking for industry experience, domain expertise, innovative technology and world class after-sale support, you can count on us to meet your needs now, and well into the future.

Find out which Thermo Scientific informatics solution is right for you. Visit **www.thermo.com/solutionfinder**.

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

We have a solution for you, no matter what industry you are in:
pharmaceuticals, food and beverage, industrial chemicals, petrochemicals, metals and mining, water and wastewater and many many more.

Part of Thermo Fisher Scientific

Thermo SCIENTIFIC

Company Index

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

1. **Title of publication:** Bio-IT World
2. **Publication Number:** 1538-5728
3. **Filing Date:** 10/1/07
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: Deborah M. Walsh (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		
* Paid/Requested:	18,672	18,664
Free Distribution by Mail:	1,949	1,967
Free Distribution Outside the Mail:	1,206	1,131
* Total Distribution:	21,827	21,762
Copies Not Distributed:	534	665
Total:	22,361	22,427
Percent Paid/Requested:	85.55%	85.76%

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).

Deborah M. Walsh, Director of Audience Development, Bio-IT World, PS Form 3526

*Due to US Postal regulations, Paid/Qualified requests for digital editions are not included in the numbers above. Average digital distribution totals 9,611 additional copies each issue.

VOLUME 6, 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.

CHI

Bio·IT World®

Indispensable Technologies Driving Discovery, Development, and Clinical Trials

EDITOR-IN-CHIEF
Kevin Davies (781) 972-1341

EXECUTIVE EDITOR
John Russell (781) 972-1342

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

ART DIRECTOR
Mark Gabrenya (781) 972-1349

WEB EDITOR
Catherine Varmazis (781) 972-1344

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

VP SALES – MID WEST, SOUTH EASTERN US, EAST COAST (NH, NJ, NY, PA, RI, VT), AND EUROPE
Jennifer McElligott (781) 972-5482
jmcelligott@healthtech.com

VP SALES – CA, WESTERN US, CANADA, PACIFIC RIM, AND EAST COAST (CT, DC, DE, MA, MD, ME)
Alan El Faye (213) 300-3886

DIRECTOR OF AUDIENCE DEVELOPMENT
Deborah Walsh (781) 972-1350

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

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

ADVERTISING OPERATIONS COORDINATOR
Kathrene B. Tiffany (781) 972-5440
ktiffany@healthtech.com

GRAPHICS AND PRODUCTION MANAGER
Ann-Marie Handy (781) 972-5493
ahandy@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

Staff can be reached by e-mail in the following form:

firstname_lastname@bio-itworld.com

editor@bio-itworld.com

250 First Avenue
Suite 300
Needham, MA 02494

You Stand in Front of Our Instruments All Day... We Stand Behind Them...24x7

High Value, Substantially Expanded Services from Applied Biosystems

With over 25 years of experience in the development, manufacture, and service of innovative instruments, and with over 1000 highly trained, dedicated service professionals, AB Global Services is uniquely qualified to deliver a full suite of real-world lab services. From Remote Services to On-Site Application Consulting to Qualification and Professional Services and everything in between, AB Global Services is your value-added partner to help boost your productivity and maximize the return on your technology investment.

- Instrument Repair and Maintenance
- Smart Services
- Qualification Services
- Professional Services
- On-Site Application Consulting
- Training

To learn more go to http://info.appliedbiosystems.com/wwservice or contact your local Applied Biosystems sales representative.

Applied Biosystems

Single Minded Sequencing

After going public earlier this year, Helicos prepares to unveil its long-awaited "DNA microscope."

BY KEVIN DAVIES

CAMBRIDGE — Four years ago, biotech entrepreneur and venture capitalist Stanley Lapidus was browsing an issue of the *Proceedings of the National Academy of Sciences* when a paper by Stephen Quake caught his eye. "It took my breath away!" Lapidus recalls enthusiastically as we sit in a conference room at Helicos' headquarters in Kendall Square, Cambridge.

"Viewed through one facet of the prism, he'd done very little. He'd gotten one strand [of DNA] to go to 6 bases. But it was addition of single bases to a single molecule. It didn't escape my notice that this was worth pursuing."

Quake's reaction when Lapidus and Flagship Ventures CEO Noubar Afeyan suggested they form a company? "Steve thought we were nuts!" says Lapidus. "He'd started Fluidigm, and Fluidigm technology was farther along than the technology that became Helicos." But in July 2003, Lapidus, Afeyan, and Eric Lander flew to Los Angeles to meet Quake and sketch out the first patents. By the end of 2003, Helicos had raised $27 million and was ready to begin operations.

Revolution

An engineer by training, Lapidus' business background is the world of diagnostics. He started a pair of companies, Cytyc Corp. and Exact Sciences. The idea at Exact was to extract DNA from stool samples and look for mutations associated with colon cancer. But such indirect methods made little sense to him. "One does indirect science because one can't do the direct experiment... I kept wondering, Why aren't [oncologists] just sequencing 1,000 tumors? Then I learned the idea of doing very high-throughput sequencing ... what [COO] Steve [Lombardi] calls the DNA microscope."

Lapidus speaks justifiably proudly of his managerial team. COO Steve Lombardi worked at Applied Biosystems (AB), building its sequencing business to $800 million annually, and was formerly at Affymetrix. Leading the R&D effort is Bill Efcavitch, credited with building much of AB's product line, including all its sequencers except for the 3730. The advisory board includes luminaries such as Lee Hood, Gene Myers, and John Quackenbush.

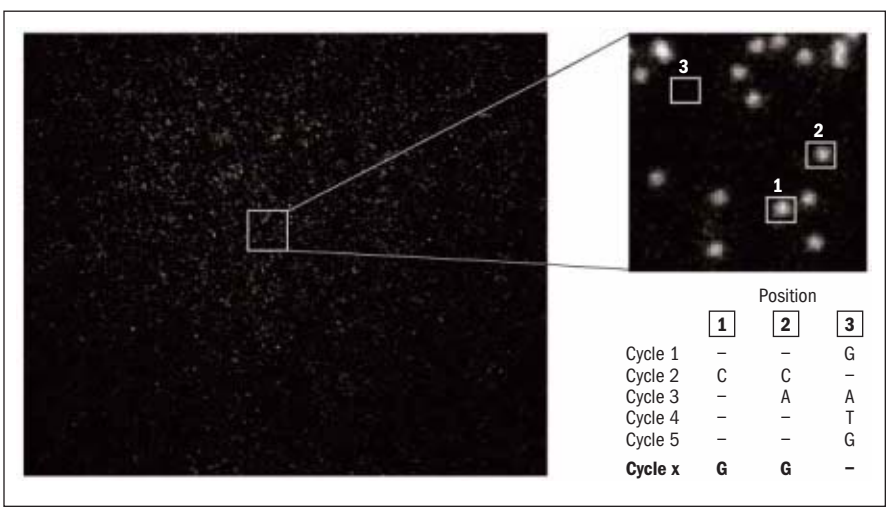
In June, Helicos raised $46 million in its IPO — below target, but "we raised the money we needed," says Lapidus. Indeed, when viewed in the context of this being the first life science tool company to go public in some six years (Lapidus cites Third Wave in 2001), it's no small achievement.

CFO Louise Mawhinney, who was formerly at ArQule, adds: "We hadn't published any papers, [investors] had to take it on faith we could actually sequence single molecules. We didn't have a single order or a commercial product. It's amazing and a great tribute to the VCs and to Stan that the market would even listen to us."

Within weeks of closing its IPO, Helicos had two prototype HeliScope instruments running, edging toward the 90 mil-

(CONTINUED ON PAGE 12)

The HeliScope will likely lead the field in price and raw sequence output.



True single molecule sequencing: 40,000 DNA molecules, imaged in less than 1/15 of a second. Right, individual base cells.

DNA Microscope

(CONTINUED FROM PAGE 10)

lion bases/hour benchmark slated for commercial release at the end of 2007. The HeliScope is a single-molecule imaging instrument that measures the fluorescence of single dye molecules. The main components are an imaging head with four lasers; a cooled CCD camera optimized for low image acquisition time; and an XY stage that moves swiftly at 15 frames/second. Because image stability is critical, the machines are housed in a 500-pound slab of granite shipped from New Hampshire.

The prototype HeliScope differs from Quake's lab-made prototype in several key respects. Quake's 2003 paper used fluorescence energy transfer (FRET), but Helicos found that FRET was limited to short read lengths and added extra cost and complexity to the instrument. Breakthroughs in surface chemistry obviated the need for FRET.

That was "a big deal for us," says Lapidus. "We had to develop surfaces that rinse well and without using ring structures in the solvents. The experiments take days, so that stability throughout a multi-day period is not easy to do." The flow cells have now been upgraded to minimize the problems caused by background fluorescence. (See "A Frightening Computational Problem," p. 14.)

HeliScope Developments

Another key development was engineering new "virtual terminators," thanks to Helicos' 7-person organic chemistry team. The virtual terminators help solve the issue of repeat sequences (homopolymers). One way to read them (as Roche/ 454 does) is to measure intensity differences. "The problem with those is you can tell 1 from 2, but 3 from 2 is only 50% as bright, 15 from 14 is very difficult. With single molecules it's even worse, because of quenching," Lapidus says.

The virtual terminators interact with DNA polymerase to slow down subsequent base incorporation. "You don't wait for the reaction to run to completion," he says. "You wait a minute or two and wash it out. The chance of a second base being added with the first base there is very, very low."

Lombardi fleshes out the details: "We look at 600,000 growing DNA strands/second [40,000 growing strands/image, moving 15x/second], 90 million bases/hour. There are two flow cells — one does chemistry, one does imaging. When [the camera] takes an image, we're at a density of 1 strand/square micron, or 100 million/cm^2. Three billion strands in both flow cells." (See "Clusters and Coordinates.")

The early access program was essentially a collaboration with members of Lee Hood's group (Hood is a Helicos SAB member). The next phase, Lombardi explains, is to select "labs that we think best represent the market opportunity, who want to be early adopters. We'll place the instrument in the labs and confirm what we've done on the manufacturing floor." Those instruments will go through a rigorous validation verification, then have a Field Service Engi-

(CONTINUED ON PAGE 14)

> "This is... Helicos' shot at being more than a footnote in science."
>
> **Stanley Lapidus**, Helicos

Clusters and Coordinates

Steve Lombardi

"We generate a lot of data," says Steve Lombardi, rather stating the obvious. "We do near real-time image processing. We take an image of an XY coordinate, we do that 45,000 times, switch the cells, do the other, then switch again.

"We're reading these images through time and connecting the dots. In real time, we're taking the entire data set and compressing it into XY space where we only have the signal. That's a pretty heavy computational process, but it keeps up with the instrument.

At the end of the run, you don't have to process the data and wait — you can start the next run while you're porting the data off to the informatics.

"It's a tower, with graphic processing chips used in the gaming industry, incredibly inexpensive and phenomenally compute intensive. We can do it for a relatively reasonable amount of money, but it's a significant part of the cost of the box."

In reference to the bioinformatics alignment engine Lombardi said, "We'll source this — an informatics tower that at today's costs is $150,000. It's a 10-multiprocessor Linux Cluster Blade that has field programmable gate arrays that can take 30 billion bases and align them to a reference sequence in less than 24 hours."

Lombardi and Kristen Stoops, the company's head of bioinformatics business development, are evaluating potential informatics collaborators. "We're going open source with the image processing software and the informatics... Then we'll be working with partners to figure out who are the best partners to source this. The Broad Institute isn't going to buy this, but the academic health centers will." **K. D.**

Biocomputing systems that analyze the environment without warming it.

High-Performance Biocomputing Solutions

The CodeQuest™ and DeCypher® systems from TimeLogic® handle next-generation genomics projects faster than most clusters, yet require less electricity than a standard light bulb.

Using the FPGA-based DeCypher Engine™ accelerator cards, you can apply the performance of hundreds of CPUs toward:

- Mapping short-read sequences, SNPs and microarray probes with speed and specificity using Tera-BLAST or Tera-Probe™
- Matching millions of nucleic sequences to nt with Tera-BLAST™
- Comparing proteins to UniProt using Smith-Waterman
- Identifying novel protein family members using HMM searches

Keep your genomics projects on track and your computing budgets above water. Contact TimeLogic for a free performance metric today!.

DeCypher®

Scalable, modular Solutions for the Enterprise.

CodeQuest™

Accelerated desk-top workstation to speed genomic workflows for your lab

North America
Active Motif, Inc.
Toll Free 877-222-9543 x 5
tlsales@activemotif.com

Europe
Active Motif Europe
Direct +32 (0)2 653 0001
tlsales@activemotif.com

Asia
Active Motif Japan
Direct +81 (0)3 5225 3638
japantech@activemotif.com

Additional international distributors: www.timelogic.com/distributors

TimeLogic®
biocomputing solutions

A brand of Active Motif,℠ Inc.

DNA Microscope

(CONTINUED FROM PAGE 12)

neer and a Field Application Scientist accompany the instrument once it is installed in the customer's lab to "confirm the same results in their lab as we got on the factory floor."

As for cost, Lapidus simply says the instrument will be "in the ballpark of mass specs."

Competitive Landscape

Helicos will debut its first instrument two years after the first machines from Roche and Solexa (now Illumina). Doesn't that set Helicos back, given its competitors have already booked orders for 200 instruments? Lombardi turns on the question with the aplomb of a clean-up hitter. From protein sequencers to oligonucleotide synthesizers, DNA sequencers to microarrays, being first to market is no guarantee of ultimate success he says. "This feels like 1997," says Lombardi. "AB had the 377, [then] up popped the MegaBACE. Oh my god, AB's dead! Darned if the 3700's arrived a year later. The MegaBACE was a major advance but it wasn't right. We think this is the right product for the market."

"It's fine we're not the first in the market," agrees Lapidus. "We believe we have the best platform and the best trajectory. I didn't start this company to do the Neanderthal [genome]... It's really about medicine. It's about reducing morbidity and reducing mortality, living longer and better lives." He adds: "It's inconceivable ten years from now that anything other than single-molecule technology will be used for nucleic acid measurements. Maybe it isn't Helicos, but I believe it will be."

As excited as he is about the technology and the business opportunity, Lapidus relishes the bigger picture. "This is my shot — and Helicos' shot — at being more than a footnote in the history of science," he says. Recalling the importance and excitement of the era of microbe hunting, Lapidus sees the coming era of next-generation sequencing as "an age whose opportunity to change the quality of life of mankind for the better is just as profound as it was then." ●

A Frightening Computational Problem

Tim Harris, Helicos' research director, is a veteran of the single-molecule sequencing world. The Bell Labs veteran joined SEQ — the first commercial single-molecule biotech firm, founded by Kevin Ulmer in 1987 — as director of research in September 1996. During a recent talk at Harvard University, Harris asked rhetorically, "So why did this [Helicos technology] work? People have been trying to do this since 1987... When Stan Lapidus called me up, and said 'Tim, I got just the job for you, when can you come to Boston?' I said, 'Stan, my friends and I have been failing at this for 15 years, are you sure?' His answer was, 'Yeah, I think it'll work this time.'"

To sequence 90 million bases/hour, Harris describes the recipe as follows: first, shear genomic DNA with DNAses into fragments of about 100 basepairs. Melt the strands and add terminal transferase to produce polyA tails, with a fluorescent tag at the end. Meanwhile, on a glass surface randomly immobilize a forest of polyT primers – about 1 per square micron. Then swish in the sample to hybridize to the slide. The "off rate" at 37 degrees C is about 1 week. "I can do a lot of sequencing [in that time]," says Harris.

According to Harris, there is 1 molecule of DNA per square micron, or one million per square millimeter. (Each spot is localized to a mere 15 nanometers.) The camera takes a picture, the dye is removed, the next base is added, and the cycle repeats 100–200 times. A 25-base read takes two days and is the median length; 20 percent of DNA strands grow longer than 30 bases. But even under the best conditions, the platform still fails to detect incorporation 1–2 percent of the time, a problem that is actively being addressed.

One of the barriers to getting single-molecule sequencing to work has been preparing a surface that can handle micromolar concentrations of dyes and then rinse the dyes out so the only 10–20 remain per 1000 DNA templates. Harris hired Mirna Jarosz ("a real smart lady") who solved this rinse failure background fluorescence issue. Harris notes two other major keys to the success of the Helicos platform: an active nucleotide analog/polymerase combo that will incorporate bases, and better purified reagents than are commercially available

Helicos has worked hard to solve the homopolymer problem. The commercial solution is to use kinetically engineered nucleotide analogues that have "no run through." Because incorporation errors are random, Harris says it's very unlikely to happen twice. But Helicos is exploring two-pass sequencing, in which a new primer would be added after the first run to the other end, allowing the strand to be sequenced in reverse direction.

Align and Repeat

Sequence alignment is "easy if the genome is small, and requires real creativity if the genome is big," says Harris. "You take 7 or 9-mer pieces of your read, you go shooting through the genome, and ask, does the rest of the sequence look like it fits there? It's a substantial computational problem."

Harris calls the informatics infrastructure "pretty frightening." The HeliScope will generate 20 terabytes of image data/day. "There's not a plan to save it, which scares me, I have to say. I've never thrown away my data in real time," Harris laughs nervously. "It generates 1 TB of actually analyzed data (read strands). It's a bit of a frightening computational problem." **K.D.**

Applied Biosystems Software Initiative for Next-Gen

Development plan enables analysis software.

BY KEVIN DAVIES

Applied Biosystems (AB) has launched an initiative to help life scientists and independent software vendors develop software applications for its SOLiD next-generation DNA sequencing platform. AB says it is expanding its software development community to include sample data sets, data file formats, and data conversion tools for the newly released SOLiD.

The SOLiD system enables "myriad applications in genotyping, sequencing, epigenomics, and gene expression," says Michael Hadjisavas, AB's director of commercial development. "One potential expectation of the software initiative in this area is to enable the development and availability of analysis software that will enable researchers to visualize and interpret data for these various applications."

> The promise of next-gen sequencing technologies will be software applications for data analysis.

AB says it is the first next-generation sequencing manufacturer to make these tools available to the bioinformatics community, and wants to help tackle the significant challenges associated with analyzing and managing the vast amounts of sequence data. Last summer, dozens of researchers from academic and commercial institutions gathered at AB's headquarters in Foster City, CA, to share best practices on various next-generation sequencing software development initiatives.

"The promise that next-generation [sequencing] technologies will generate better data, faster, and at a lower cost will only be realized when there are sufficient software applications that allow researchers to analyze this data," said Michael Wittig, a bioinformatician at the Institute for Clinical Molecular Biology in Kiel.

A key issue, according to Darren Platt, head of informatics at the Joint Genome Institute (JGI), is not just the "the quality of the data each platform generates, but ... how willing each vendor is to collaborate with the research community to provide the necessary tools and resources for developing the required software applications." AB's willingness to share information about the SOLiD system "will enable us to rapidly develop the software applications our laboratories need to accelerate next-generation research projects," says Platt.

Another goal of the software initiative, says Hadjisavas, is "to enable the development and availability of analysis software that is independent of application and viewed as foundational to all types of analysis. Examples of such foundational analyses include software for alignment, assembly, annotation, quantitation, management, and visualization."

Hadjisavas expects the initiative "will alleviate the potential bottlenecks that researchers will encounter that are related to analysis, visualization, management, and interpretation of data. He says AB is "proactively engaging" in discussions with customers and vendors to develop software requirements that will stimulate the development and availability of software tools for this "compelling technology."

AB is looking to aid a bioinformatics community to further application development in a variety of research areas, including whole genome sequencing, chromatin immunoprecipitation (ChIP), microbial sequencing, gene expression, microRNA discovery, digital karyotyping, and epigenetics. ●

WEB 2.0 TOOLKIT

Nextrials has just released Prism Tools, the pharmaceutical industry's first intuitive Web 2.0 toolkit that leverages the **Clinical Data Interchange Standards Consortium**'s (CDISC) Operational Data Model standard to deliver more useful data through a user-friendly, dashboard-oriented experience. The user interface for Prism, Nextrials' clinical trial management and data collection software, Prism Tools assembles information in multimedia formats using Flash and other graphical applications.

STATE TECH GRANT

Illinois governor Rod Blagojevich announced a $100,000 grant to the **Chicago Technology Park** as part of the state's Biotech Training Investment Program. The grant will fund the BiTmaP Program, a bioinformatics training curriculum originally started by a U.S. Department of Labor grant that helps IT workers make the transition to life sciences.

AFFYMETRIX-APPROVED

GenoLogics' Geneus product has achieved GeneChip-compatible status with **Affymetrix**, and the company has joined the GeneChip-compatible Applications Program. Geneus, a LIMS and scientific data management system, is now seamlessly integrated with the Affymetrix GeneChip platforms, GeneChip Operating Software, and GeneChip Command Console Software.

COLLABORATION TERMINATION

Alnylam and **Merck** have ended an ongoing RNAi drug-development alliance worth more than $120 million. The companies mutually agreed to terminate their July 2006 amended and restated agreement. As a result, Alnylam has rescinded all grants of its intellectual property related to current and future Merck development programs, including the partnership's former co-development programs.

Metagenomics On the Case

454 finds clues to the mystery of colony collapse disorder.

BY KEVIN DAVIES

In important research published in *Science*, 454 Life Sciences (Roche) says its technology has provided a critical clue into the perplexing mystery of the decimated bee populations in North America. The so-called colony collapse disorder (CCD) has mystified biologists for over a year, with an estimated 10 billion bees (2.4 million hives) lost. First recognized last year in the Unites States, similar outbreaks have since been reported across continental Europe.

454 and scientists at Columbia University led by Diana Cox-Foster and Ian Lipkin undertook a "metagenomic" approach to identify possible culprits. The results revealed traces of several foreign viruses, fungi, and bacteria associated with the honeybee affected populations. However, one sequence appears particularly significant: that of the Israeli Acute Paralysis Virus (IAPV), which was only detected in the collapsed colonies.

The authors of the study stress that this does not prove a causal relationship between IAPV and CCD. "Nonetheless, the prevalence of IAPV sequences in CCD operations, as well as the temporal and geographic overlap of CCD and importation of IAPV infected bees, indicate that IAPV is a significant marker for CCD," the authors conclude.

Adds Lipkin: "We view this work as a model for investigating epidemics of unexplained infectious disease." ●

PAUL J. RICHARDS/AFP/GETTY IMAGES

Sequencing technology allowed researchers to compare CCD and non-CCD populations.

Further Reading: Cox-Foster, DL *et al*. "A metagenomic survey of microbes in honey bee colony collapse disorder," Science, September 7, 2007.

Connecting Neuroscience Knowledge

ERIC K. NEUMANN

Neuroscience involves many sub-disciplines: cell biology, electrophysiology, molecular genetics, chemistry, endocrinology, pathology, pharmacology, imaging, computer science, and so on. In recent years, major advances have been made by coupling knowledge across multiple research areas — Alzheimer's, depression, schizophrenia, etc. — all enabled by the exchange of data and analyses. The surge in information growth makes reliance on information systems to connect and access relevant items essential.

Information capture and analysis have always been key ingredients in neuroscience — even more so now with tools such as functional MRI for studying brain activity. The rise in integrative informatics is the reason NIH funded the BIRN (Biomedical Information Research Network) project. Conceived in 2001 to help connect large-scale biomedical informatics collaborations, BIRN supports projects that focus on neuroscience at multiple levels (e.g. Morphometry BIRN, Function BIRN, Mouse BIRN). The ability to share analytical tools and data is central to BIRN, and sets a course for the future of all biomedical research.

The term "cyberinfrastructure" was coined in 2003 by an NSF committee, recognizing the need for new mechanisms of information handling and exchange. Data repositories and computational tools need to use the Internet, but currently this is done through web pages that may include manually driven database interfaces. This requires substantial effort to access even the simplest data sets, and only a handful of machine-to-machine data transaction systems exist (for example, Web services).

Today, much of what is being created for systems integration requires the inclusion of semantics. Terminologies must be concisely defined and logically inter-related in areas such as anatomy, molecular biology, diseases, neurochemistry, and others. Without definitions, no amount of computational power can untangle different data descriptions created by different researchers. Consequently, most of the above projects involve ontology development and management. Such ontologies need to be 1) used by all research groups regardless of their locations, and 2) defined in such a way that they can be combined as necessary by inter-disciplinary projects. Both requirements are addressed by Semantic Web standards.

Working with university members of BIRN at Yale, Stanford, Tennessee, San Diego, and Drexel, the World Wide Web Consortium (W3C) Health Care and Life Sciences Interest Group has assimilated many forms of neuroscientific data and structured them using RDF and OWL. This aggregate of neuroscientific knowledge was part of a demo presented at WWW2007 (Banff, Canada) and ISMB 2007 (Vienna). Data sources currently include: BrainPharm, Pubmed, Entrez-Gene, Uniprot, MESH, BAMS, Reactome, Gene Ontology, Allen Brain Atlas, NeuroCommons Annotations, NeuronDB, AlzGene, SWAN, MammalianPhenotype, Pubchem, and Homologene.

Text Mining Research

NeuroCommons, a project within Science Commons at MIT, is using text mining to extract neuro-molecular relations from text, representing them as RDF. BrainPharm is a data resource from Yale that supports research on drugs for neurological disorders. The Allen Brain Atlas has assembled multiple gene-probed slices of mouse brain. (See "Allen Brain Atlas Updated," *Bio•IT World*, September 2007) And SWAN is an NIH-funded project that allows scientists to directly annotate knowledge onto findings using RDF.

The demo user interface consists of a SPARQL query page that permits a wide variety of questions regarding genes, neurological diseases, neuroanatomy, and publications. Examples include:

- Find all publications with *neural dendrites* in their description.
- Show all genes expressed in brain region CA1 involved in signal transduction.
- Find all papers on Parkinson's Disease that involve gene products localized in the nucleus

Results can be formatted as tables or even as RDF graphs. As RDF, additional tools can process the data for enhanced scientific views. Moreover, tools such as Google Maps can also be applied to the output from a query.

The future of cyberinfrastructure for biomedical research is becoming a reality: a connected research community more effectively utilizing data and computational resources from different areas. Providing a new infrastructure that will connect different forms of knowledge is an essential element of biomedical research (see NIH's KEBR: http://esi-bethesda.com/ncrrworkshops/kebr/index.aspx), and in the future, these resources will be used by researchers in R&D and health care.

Eric K. Neumann is the Director, Clinical Semantics Group; MIT Fellow, Science Commons; and co-chair, W3C Health Care and Life Sciences Interest Group. E-mail: eneumann@alum.mit.edu

Next-Generation Sequencing Solutions

BY WILLIAM VAN ETTEN

"All that you are, is either protein or the result of protein action." — J. Craig Venter

I'm not sure whether Venter said this first, but I heard it first from him, and it's still one of my favorites quotes. The deeper meaning is this: the uniqueness of an individual can be reduced to their unique set of proteins, encoded within their genome, and is located within the estimated 0.1 to 0.5% differences between any two genomes. The Human Genome Project identified most of these differences, but it leaves me wanting to know "my" differences, my "personal genome." And I'm not alone in this view. This year, we've witnessed the personal genome sequencing of both Venter and James Watson, and startups that hope to capitalize on the business of personal genomics, such as Navigenics (See p . 6) and 23andMe.

Sequencing the first human genome required ten years at a cost of around $1 billion. Much of this time and money was devoted to preliminary science that doesn't need to be repeated in order to sequence subsequent human genomes. Hence sequencing another individual human genome using conventional Sanger sequencers might require 10,000 instrument days (30 instruments for 1 year) for around $10 million. Venter's personal genome was sequenced in this way. He sequenced his diploid genome (requiring double the effort), but borrowed some of his sequence from the Celera genome project, so this probably required something less than 60 instrument years and $20 million. In contrast, Jim Watson's genome was sequenced by 454 in two months for $2 million. The cost of a personal genome should drop below $100,000 within a year.

Each next-generation sequencing vendor has its niche (e.g longer reads, greater accuracy, greater data density, etc), but they all offer 100 to 1000-fold improvements in throughput and cost over Sanger sequencing. How will the ability to generate so much more sequence so cheaply affect the sequencing market, and what are the IT implications for analyzing all this data?

Overall, the market trend is a lot more sequencing being performed by many more, smaller research groups. Assays previously performed using mass specs and microarrays are now being performed using next-gen sequencing. Beyond individual reference genomes, we're resequencing intra-species variants for phylogenetic, association, and population studies. Clinical researchers are thinking about how to include a personal genome within a patient's electronic medical record, permitting a physician's diagnosis and annotation.

IT Implications

The BioTeam has had the opportunity to interact in varying degrees with each of the next-gen sequencing vendors, and with several researchers utilizing the vendors' instruments. The storage and computing requirements for each vendor's instrument are unique, but can be generalized to 0.2-6 terabytes (TB) of primary image data generated daily and a minimum of 4-8 CPUs in continuous operation to perform image analysis, base calling, and sequence assembly in order to keep up with data production.

The vendors' solution to this problem is to recommend (or include with their product) a Windows or Linux workstation to drive the instrument with a large local disk to collect the image data. These data are subsequently copied to a shared network file server for analysis by a small (~4 node) UNIX cluster. The advantage of a small cluster with large network storage is that it is a well-established architecture for providing many processors working together to analyze a common set of data. The disadvantages are that this particular problem is more IO-intensive than CPU-intensive (a bottle-neck for clusters) and that even though clustering has become easier over the years, configuring and maintaining the dependent shared network services is still beyond the abilities or duties of a biologists and hospital technicians.

Overall, the market trend is a lot more sequencing being performed by many more, smaller research groups.

We have participated in the deployment of hundreds of UNIX clusters for many different compute-intensive research problems, helping to make clustering easier for researchers who don't know how (or don't want to know). But is a small cluster and shared network file server the most suitable solution for this particular problem? Rather than simply throw the solution we know at the problem, we re-examine the requirements and propose a better-suited alternative.

User: The average biologist or hospital technician knows their research, knows computing from the perspective of desktop productivity tools, but does not have IT experience. Small research labs probably don't have support of dedicated IT staff.

Processing: Image analysis, base calling and sequence assembly require between 4 and 8 of today's modern processors in continuous operation to keep up with data production.

Storage: Moving terabytes of image data from one disk to another over a gigabit TCP/IP network using a modest network file server takes hours, even days. Moving that much data through a direct connection on the same computer takes minutes to hours.

Data Archival: Although disks are cheap, DNA sequencing is cheaper. It might cost less to sequence a genome again than to archive the primary data from sequencing it the first time.

Data Management: The results generated by next-gen sequencers are massive. In order to become useful information, these results must be related and re-related by researchers to answer whatever questions they posed by the experiment. This is clearly the most challenging and unaddressed issue.

Alternative Solution

The ideal compute and storage engine to support a next-gen sequencer is: 1) easy enough for a biologist or hospital technician to operate and manage; 2) provides 4 to 8 processors; 3) provides 10-20 TB of fast disk; 4) and the disk presents itself directly to both the instrument driving workstation and the compute engine without copying the data.

So we thought... let's replace the 4-node cluster with a single 8-core system. This provides the necessary compute power, but gets rid of all that clustering business. Then replace the shared network file server with a direct-attached fiber channel disk array. This provides the same storage capacity but much faster IO performance. Finally, move the independent Windows/Linux instrument driving system to a virtual machine operating on the same system. This permits the instrument to write to the same disk that the compute engine reads from without copying the data. The disadvantages are that we can't easily add more processing power if we want to, and the CPU/IO of data analysis might hinder data gathering in the VM layer.

For our validation experiment, we used an Apple Mac Pro with dual 3-GHz Quad-core Intel Xeon processors and 8 GB of DDR2 RAM, operating Mac OS X 10.4, attached by Quad-Channel 4-Gb Fiber Channel to 10.5-TB Apple Xserve RAID configured with RAID level 5 mirroring. With this hardware configuration, we used Illumina's data analysis pipeline (version 0.2.2.5) to analyze a small run consisting of a sequenced bacterial artificial chromosome including 68,000 images. Under these conditions, the analysis execution time was around 4 hours — the same as a small cluster with network disk. We captured a disk image of Illumina's Galaxie workstation and demonstrated that we could run it within a Parallels VM layer. (We did not attempt generating new sequence while analyzing prior sequence.)

Over the next few years, a lot more sequencing will be done by smaller groups that won't have dedicated IT staff to manage clusters and file servers. A single 8-core system with fast, direct disk may be adequate and perhaps better suited to the problem.

Email William Van Etten at bill@bioteam.net.

Wyeth's Approach to Clinical Biomarkers

MALORYE A. ALLISON

Most companies have the same reasons for doing biomarker research. "We have to understand how a drug is working before we invest large sums of money into Phase II and III trials," says J. Lynn Rutkowski, co-leader of clinical translational medicine at Wyeth Research. "If we can work through the mechanism, we can have much more confidence advancing the project."

That's the argument for using biomarkers, but how does that actually translate into specific programs? According to Rutkowski, it all begins early in development. "You need a strategy in place so you have time to do the research you need to fill in gaps and get biomarkers you have confidence in."

Every compound at Wyeth — the company currently has 75 to 80 compounds in development — is evaluated for potential biomarkers. Rutkowski is responsible for the clinical development side of Wyeth's translational medicine department. The team overseeing each compound includes a discovery biologist, clinical pharmacologist, and translational scientist. "There's a dedicated person who is responsible for developing and delivering response biomarkers," she says.

Last year, Wyeth initiated a collaboration with four Scottish universities (Aberdeen, Dundee, Edinburgh, and Glasgow) called the Translational Medicine Research Collaboration (TMRC). Wyeth is investing almost $86 million for biomarker discovery, including ten in cardiovascular and metabolic disease, four in inflammation, and seven in neuroscience.

"For each target, you want at least one biomarker unique to that target," Rutkowski explains. Wyeth is continually developing new markers, which may take many forms. In stroke, for example, in addition to imaging technology, Wyeth is using rehabilitation tools to gauge patients' responses. "We are taking advantage of robotic instrumentation for therapy that can also provide a quantitative assessment of motor-function recovery," she says.

> "There are so many technologies emerging. The moment you commit to one, there is another right behind it."
>
> **J. Lynn Rutkowski**, Wyeth Research

Alzheimer's Advances

Alzheimer's disease (AD) is a particularly important indication at Wyeth, which has 11 AD compounds in development. The company's long-term strategy involves molecular markers, structural and functional brain imaging, and physiological, behavioral, and associative learning tests.

One problem is that it takes so long for AD patients to either show disease progression or improvement. And as it's not possible to take samples of brain, molecular markers are picked out of cerebrospinal fluid or plasma. "If you are measuring something in a surrogate fluid, you still do not know exactly what is happening in the target tissue," Rutkowski explains. However, the growth in biomarkers for AD is revealing new insights about the disease.

"We know there are familial forms of Alzheimer's, and we know that mutations in gamma- and beta-secretase cause a build up in plaque," Rutkowski says. Another hot target in AD are the neurofibrillary tangles seen in patients' brains after autopsy. Plenty of markers are used to track the various states of molecules that play a role in plaque and tangle formation.

"Ideally, for a brain disorder, we would like to have a receptor ligand that could be the signal of some crucial event," says Rutkowski. In concert with imaging, this would be a very specific approach. "We could verify that the drug gets to its target."

Wyeth scientists are pursuing another type of plaque as an exciting new target — the so-called "vulnerable" plaque found in peripheral arteries that can rupture and cause heart attack or stroke. "Many of these cardiovascular diseases are linked," Rutkowski says. "If you have lots of plaque, for example, you are at risk not only for ruptures; the vessels may get stiffer, causing high blood pressure." Wyeth has several drugs that are targeting different processes, including inflammation.

Wyeth researchers are particularly interested in FDG-PET studies that will help monitor metabolic activity in these lesions without requiring biopsies, and are exploring different platform possibilities. The company's scientists are also examining different ways of labeling macrophages to track their infiltration of the lesions. "We want these markers implemented in Phase I, and we already have a compound that is getting close to an IND," Rutkowski says.

"At this point, we can do a Phase 0, which is like a Phase Ia/Ib/IIb type of study, something that allows more exploratory kind of research with additional add-on studies," she says. With those completed, they should have some markers in hand, and feel confident that they are relevant to this condition.

Overall, Rutkowski doesn't see many dominant approaches. "There are so many technologies emerging," she says. "The moment you commit to one, there is another right behind it."

Computational Solutions for Managing the Life Sciences Data Glut

SGI has developed targeted solutions in collaboration with leading ISV's in the life sciences space, with the aim to present a complete end-to-end solution for the pharmaceutical and biotechnology research process. SGI Life Sciences solutions consist of high performance (shared memory), high throughput (Clusters) and specialized computing (FPGA's) platforms based on Linux and Microsoft® CCS operating systems. These computing platforms along with storage products have been configured to optimally address specific genomics, proteomics, cheminformatics, and other drug discovery bottlenecks within the pharmaceutical and biotechnology research workflow. SGI is a leader in high-performance computing.

Viewers will learn about:

- Targeted solutions for accelerating genomics and proteomics workflows
- Cost effective data storage solutions
- Eco-friendly, "green" computing products
- Solutions for chemistry workflows in pharma and biotech research

Bio-IT World
OnDemand Webcast

Available NOW!

www.bio-itworld.com/webcasts/sgi

Featured Speakers:

Dr. Deepak Thakkar
BioScience Segment Manager
SGI

Dr. Kevin Davies
Editor-in-Chief
Bio-IT World

sgi® INNOVATION FOR RESULTS™

www.bio-itworld.com/webcasts/sgi

Computational Biology

The Pathway is the Target

Avalon Pharmaceuticals focuses on signatures and scale in drug discovery.

BY JOHN RUSSELL

What if instead of worrying about hitting a specific target to modulate a pathway, you skipped the target part, and aimed directly at the pathway? In essence that's what seven-year-old Avalon Pharmaceuticals does. Instead of screening compounds against individual targets, it screens them against pathway signatures. Mimic the desirable pathway signature and you have a hit.

Actually, there's a fair bit more to it than that. Avalon, based in Germantown, Maryland, spent five years and $50-million building a proprietary experimental platform — AvalonRX — to perform high-throughput cell assays and transcriptional profiling. In theory the approach is straightforward and being used by others: treat cells with compounds; collect RNA; and perform qPCR-based transcriptional analysis to identify compounds that reproduce desirable pathway signatures.

The secret sauce is scale, says CEO Kenneth Carter. "There are companies, BioSeek being one, that in one form or another have created a signature database. If you give them one compound or 10 compounds or 30 compounds, they can run it in some sort of a cellular system and compare it to what they have in their database and give you some information about how those compounds are working."

Citing Avalon's collaboration with Novartis, Carters says, "The very big difference is the ability to screen 215,000 compounds [for Novartis] and look for signatures you have determined proactively to be signatures that you're trying to generate with a new drug candidate. This isn't about taking somebody's single compound or maybe 20 or 30 compounds and profiling them; it's about screening for new drugs where people have been unsuccessful in identifying drug candidates."

Cost reduction is another critical key. "If you wanted to look at, say, even as few as 10 or 20 genes, you're talking about $100 to $200 per compound that you're going to try and screen. What we did over about five years is to develop a system in which we made a large number of proprietary modifications, which we hold as trade secrets, in which we have the ability at literally less than fractions of a penny to gain the same information for a large number of genes," says the Avalon CEO.

Carter says the secret to Avalon's success is scale.

"So we don't look at the whole genome, but we use the whole genome to create a signature. We do studies on the whole genome, we reduce that to a smaller number of genes which can vary, and then we use this proprietary system that costs fractions of what it would cost the average company to start trying to screen 100 or 1,000 or 10,000 or 100,000 or a million compounds. So that's what the real proprietary advantage is."

Pipeline Progress

Pursuing this strategy, Avalon mounted an IPO, has raised more than $135 million, moved two cancer programs into early trials, and struck collaborations with Merck, Novartis, MedImmune, ChemDev, and Medarex. The company's impressive scientific advisory board is chaired by Todd Golub, respected cancer microarray pioneer and founding director of the cancer program at the Broad Institute.

Avalon's most advanced candidate, AVN944, is an IMPHD (inositol-monophosphate dehydrogenase) inhibitor for hematologic and solid tumor cancers. IMPDH catalyzes a key rate limiting step in de novo purine biosynthesis. AVN944 is reported to be progressing well in a large Phase I clinical trial for the treatment of hematological malignancies and a Phase II trial for the treatment of pancreatic cancer. Phase II trials for AVN944 in hematological malignancies are expected to start in 2008.

Avalon's pipeline also has a program targeting the Beta Catenin pathway, another well-studied pathway that's difficult to target. Compounds in this program are undergoing optimization with selection of a lead candidate expected in the coming year. Other pathways being investigated include Aurora/Centrosome, Myc, and Survivin pathways.

"In the beginning we started out to do a combination of going after already well-established but very difficult to address pathways and identifying completely new pathways. We have shifted now because as it turns out, there is so much rich, fertile soil, if you will, in pathways that are already very well validated, but difficult to drug," says Carter.

Carter has high hopes for his biomarker-based discovery engine. "It would be surprising to me if there is a database of signatures for known cancer drugs and known mechanisms of action that is broader than the AvalonRx database. I can't emphasize this enough; it is not our business model either to reposition known drugs or to do fee-for-service activities where we take somebody's drug or compound and tell them some more information. It is our business model to discover new first-in-class drugs, against pathways that everybody agrees are very important."

Still, Avalon hasn't shied away from

collaborations. A Merck deal, signed this spring, includes $200 million in milestones, "very substantial" royalties, plus the ability for Avalon to move forward on any program that Merck does not. Avalon will screen compounds from Merck's library to identify and develop inhibitors against an undisclosed target in oncology. Avalon is responsible for the selection of compound families and optimization of those compounds to a preclinical candidate stage. Merck will handle clinical development, regulatory approval, and commercialization.

In July 2006, Avalon struck a drug discovery, development, and commercialization agreement with ChemDiv for small molecule oncology therapeutics. The goal is to discover new active compounds in screens against selected targets and target pathways, which have historically been considered "undruggable." ChemDiv will provide its discovery outsource services platform, as well as medicinal and synthetic chemistry for the development of new active compounds.

Of the partnerships, Carter says, "Merck can bring a tremendous amount to us in terms of helping us learn the development process and so forth. ChemDiv is the largest source of commercial chemistry libraries. They've got a 2 million compound library or so, and they're going to bring both reagents in the way of existing chemistry and also chemistry support all the way through development process. We're going to bring all the biology and some of our development expertise and then Avalon will take the lead in actually developing the products out of those."

The MedImmune collaboration, begun in June 2005, sounds a little more industry standard. Avalon is to identify lead compounds for the discovery of small molecule therapeutics against inflammatory diseases. MedImmune is responsible for preclinical and clinical testing of resulting candidates, and all future development, sales, and marketing activities.

Of Avalon's four major partnerships, three have substantial downstream opportunity, says Carter. The fourth, with Novartis, is "a very large pilot study in which essentially they paid research support in near-term milestones to the tune of $2.5 to $3 million for a very large experiment to make sure that the technology we have would be able to do some things that they want to consider for a much broader scale potential partnership," he says.

Publicly Hopeful

Given the high capacity of the Avalon platform, you'd think the temptation to add fee-for-service offerings would be strong. Indeed, Carter agrees that Avalon has the capacity "to do a lot more screening and unique development work." Nevertheless, he insists, "it would be hard to imagine us ever having more than five or six partners. Frankly, it'll probably end up being fewer than that, but it'll be hopefully two or three partners in which we have a very deep and very intellectual property and product-rich partnership."

So far, investors seem patient. A recent analyst call lasted 28 minutes with barely a snarl. Avalon raised $20 million in a private placement last spring, presumably to help fund its ramping clinical trial activity. Its stock had a rocky first year. The stock price plunged from $10.50 at the September 2005 IPO to under the $3 by September of 2006. It has since climbed back somewhat and was recently trading between $5 and $6.

Carter says, "I think way too much is made of the hassles of being a public company. The opportunities both in terms of the stability the public markets bring at the end of the day, and the ability to create stable financial structures are many times greater in the public markets than they are in the private markets. It also brings a level of sophistication and respect that puts you in a much stronger position in partnering discussions with the larger companies.

"So yeah, the stock is doing fine. We just raised pretty significant capital, and there seems to be a lot of enthusiasm among the major institutional investors for the company," he says. "I think people understand that the development process takes awhile. I think one thing we have going for us among many of the more platform-oriented companies is that even though we've developed a unique proprietary technology, from the beginning we have also understood that at the end of the day, the cash flows from a very successful therapeutic product." ●

Computational Biology

Data Warehousing Project for the NIH

Northrop Grumman curates two data warehouses for NIAID.

BY ALLISON PROFFITT

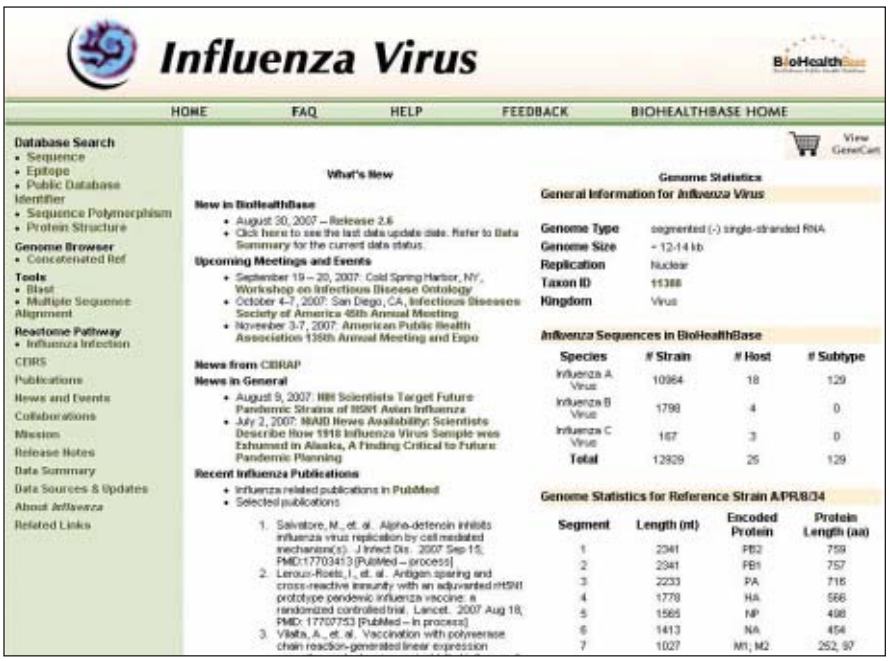
Three years into its bioinformatics practice within its life sciences division, Northrop Grumman is working on two data warehousing projects valued at over $47 million for the National Institute of Allergy and Infectious Diseases (NIAID).

"There are similarities between the two engagements," says Kevin Biersack, bioinformatics program manager for Northrop Grumman. Both data warehousing projects offer one-stop shopping to users and both make use of public data. But the projects' user communities are different.

For NIAID's Bioinformatics Resource Center (BRC) for Biodefense and Emerging/Re-emerging Infectious Diseases project, Northrop Grumman developed the BioHealthBase (BHB) system, an integrated source of complex, high-quality genomic, proteomic, and supporting scientific data. Information stored here focuses on microorganisms and pathogens. For NIAID's Bioinformatics Integration Support Contract (BISC) project, Northrop Grumman developed the Immunology Database and Analysis Portal (ImmPort) system. ImmPort houses data collected by NIAID's Division of Allergy, Immunology, and Transplantation.

BioHealthBase is open to both researchers and the public. Working with a science partner at the University of Texas Southwestern Medical Center and two subcontractors, Northrop Grumman has developed BHB to include organisms with public health and biodefense implications including tuberculosis and influenza. Biersack says that the warehouse is public resource useful for scientific research in support of vaccine development and drug discovery.

A major goal of the project is to support researchers developing rapid, inexpensive, and broad-based diagnostic approaches using genomics and proteomics.



BioHealthBase catalogues the latest research on influenza.

From the BHB website (www.biohealthbase.org), searchers can run queries, analyze their findings, and display them visually without even entering an email address.

Open Source Architecture

BHB data are culled from several public sources including National Center for Biotechnology Information databases, GenBank, UniProtKB, and internal sources. "We have firewalls, of course," Biersack says, to protect the data sources. Los Alamos National Laboratory, for instance, is currently collaborating with the BHB team to integrate their data and move their public influenza site to BHB.

Northrop Grumman curates the data as well. "We add richness," Biersack says, "by annotating entries, eliminating redundancy, and filling in missing information." Data are added and updated to the warehouse via scheduled monthly data loads.

The ImmPort system on the other hand, Biersack says, is different because access is limited to researchers funded by NIAID. "In the future, the public data will be moved 'out front,'" Biersack says, but for now, ImmPort is a semi-public warehouse.

ImmPort serves as an archive for research results for allergy, immunology and transplantation projects supported by NIAID. Researchers have access to private storage, as well as the ability to compare their data, if they wish, with other public research data based on the NIH data-sharing policy. "It's results-oriented storage," Biersack says, and ImmPort currently boasts terabytes of total storage space.

The data warehouses are web-enabled and browser based, with quarterly software updates, and the use of "mostly open source," software Biersack says. ImmPort uses Oracle, Linux, Java 2 Enterprise Edition, and Hibernate. Most of the visualization and analytical tools are also open source and have been leveraged from previous NIH-funded grants.

Northrop Grumman's contracts for the BHB and ImmPort projects expire in 2009 and 2010, respectively, and Biersack says he "anticipates competition" for these renewals. But for now, he's focused on providing new software functions to support the needs of the user communities, updating the data, adding storage capability, and "enhancing the scientific discovery process through data integration." ●

Wherever in the world you compete,
Michigan can give you the upper hand.

286
IN A SERIES OF THOUSANDS

Accuri is about to change cell analysis around the globe. Michigan is helping them extend their reach.

Jennifer Baird
President and CEO
Accuri Cytometers, Inc.

Researching the life sciences requires passion. But it's also a technology race. In Michigan, we're out front and out to win. Through some very advanced and creative thinking, Accuri Cytometers, Inc. is helping advance cellular research. Jennifer Baird and her team redesigned a technology originally intended for the automotive industry, and created a breakthrough cell analysis device. Developed with the University of Michigan, this groundbreaking technology provides life scientists with the ability to achieve greater results and do it at a fraction of the normal cost. With financial incentives from the Michigan Economic Development Corporation, this cost-effective, high-performance instrument will help labs around the world identify diseases and save lives.

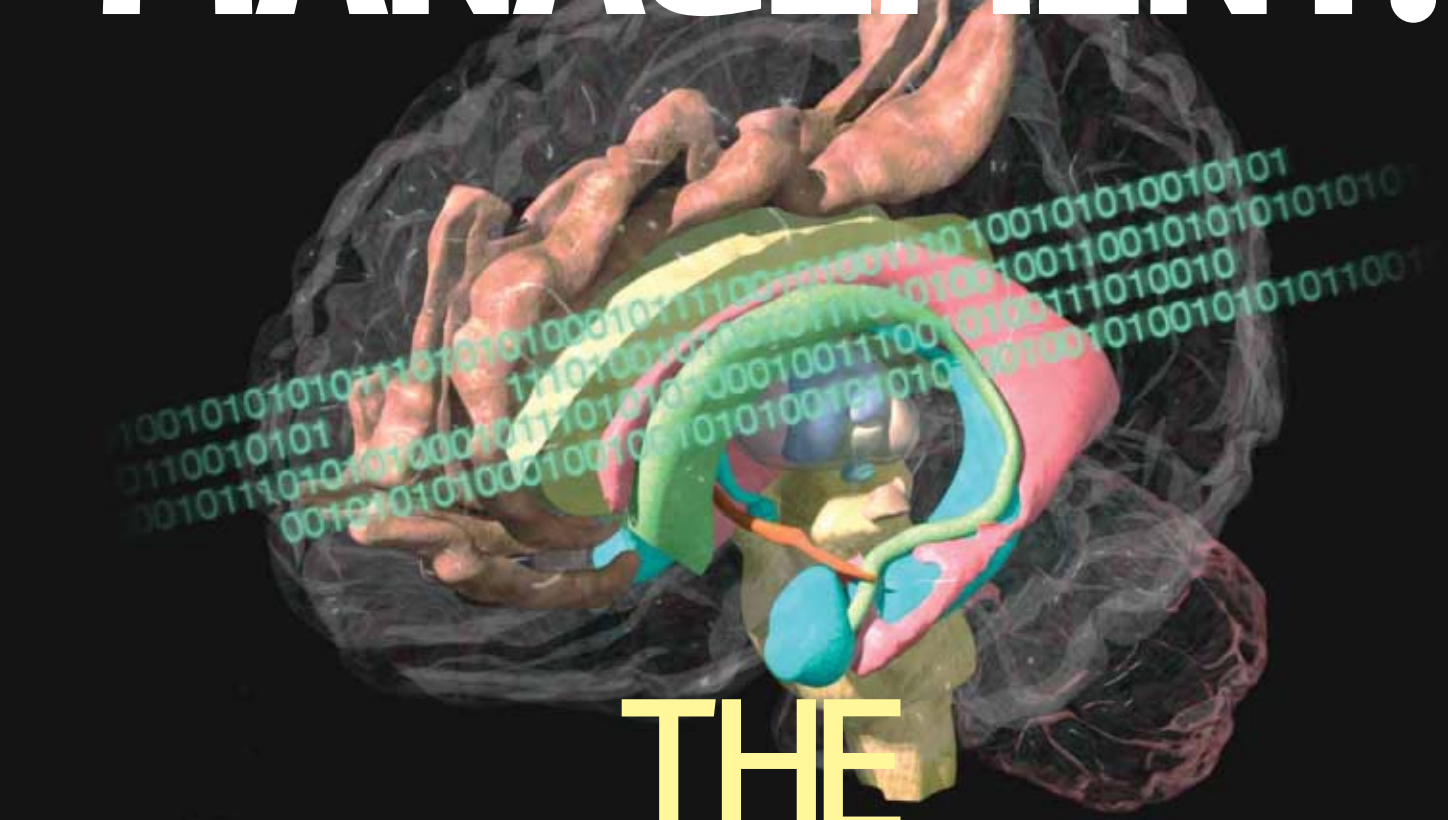
If you're looking for a place to develop your life sciences idea, we have a place for you. A place where you'll love to live 12 months a year. Where world-renowned research universities and a robust scientific community can help you win your race. **Go to michigan.org/upperhand right now.** Because wherever in the world you compete, Michigan can give you the upper hand.

MICHIGAN℠
ECONOMIC DEVELOPMENT CORPORATION

THE UPPER HAND
michigan.org/upperhand

Cover Story

DATA MANAGEMENT: THE NEXT GENERATION



New technologies aim to simplify data management while meeting life sciences' soaring computational performance requirements.

By Salvatore Salamone

Each new generation of DNA sequencers, mass spectrometers, microscopes, and other lab equipment produces a richer, more detailed set of data. We're already way beyond gigabytes (GB): a single next-generation sequencing experiment can produce terabytes (TB) of data in a single run. As a result, any organization running hundreds of routine experiments a month or year, or trying to handle the output of next-generation sequence instruments, quickly finds itself with a massive data management problem.

"Experimental data is growing exponentially," says Rico Magsipoc, chief technology officer at UCLA's Laboratory of Neuro Imaging (LONI). "With MRIs, as resolution has increased, data sets have grown in size and can be from 78 to 500 GB." According to Illumina, a fully tasked Genome Analyzer instrument yields more than 100 TB in image and processed data in a year, forcing some groups to discard raw image files to conserve storage.

In addition to lab equipment producing larger output files, another problem is that more and more of today's datasets have meta-data associated with them. "We're already seeing some adoption of Web 2.0 type technology in microscopy and microarray imaging where users are tagging, characterizing, and indexing the data," says James Reaney, director, research markets at BlueArc. This has profound implications for storage system capacity.

For many years, the way to handle data growth was to simply throw raw storage capacity at the problem. But that approach no longer works. Besides dealing with capacity challenges, life sciences organizations must also deal with performance, management, and energy issues when it comes to their storage systems.

Performance Matters

Lab data needs to be processed, analyzed, and visualized to be of any value. Typically, this requires the use of high performance computing (HPC) clusters whose nodes must be constantly fed data. Moving the data on and off of storage devices to the cluster nodes becomes the challenge.

This was an issue at the LONI. The lab's researchers and their collaborators conduct research into Alzheimer's, autism, schizophrenia, and other diseases. The lab scans humans, primates, and rodents producing 2D MRI slices of the brain. These scans are then concatenated into a 3D model.

"Four years ago, we had a SAN [storage area network] and two front-end servers that fed 20 or so machines," said Magsipoc. "We then got a 300-plus node cluster and the CPUs would be data starved."

This is a common problem in the life sciences. Many organizations moved from high-end mini-computer systems to clusters at the same time the volumes of data that needed to be analyzed were skyrocketing. This situation creates a bottleneck. Many storage systems simply cannot keep data flowing fast enough to keep computations rolling along. Magsipoc evaluated a number of systems, opting for one from Isilon that addresses the performance issue while offering fast provisioning of new storage capacity so that additional storage could easily be added.

Keeping Pace with New Data Trends

Life sciences organizations typically have masses of data that are never modified after they are initially created. A prime example of these data are the files generated by a lab experiment. Certainly, those data are often analyzed and visualized, but the original data are not changed.

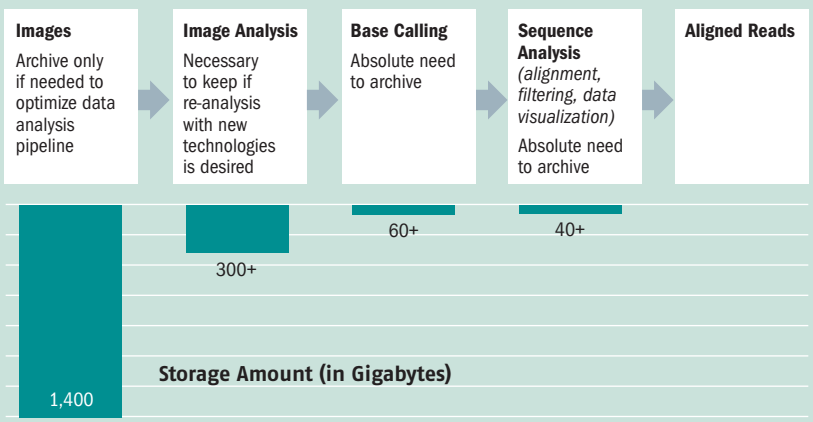
When an experiment is run, the data need to be stored on system that has the appropriate performance capabilities to support whatever analytic or visualization workflows are used to process the information. Decisions must quickly be made as to how to cost-effectively manage such reference or archival data.

"Ideally, researchers want to keep all of the [processed] data on disk," says Magsipoc. But this is not practical given the volumes of data many organizations are dealing with today. The issue is not unique to the life sciences, of course. Many industries such as financial services, oil and gas, and manufacturing must deal with data growth, archival storage issues, and matching storage performance to computational requirements.

In fact, most companies are finding they simply need to keep data for longer

Next-Generation Data Analysis Lifecycle

Schematic shows various tiers and data quantities produced by Illumina's 1G Genome Analyzer per typical run.



SOURCE: Illumina, Inc. and Blue Arc Corp.

COURTESY, ARTHUR W. TOGA, LABORATORY OF NEURO IMAGING AT UCLA

periods of time. Some industries do so to meet regulatory requirements for data retention. In the life sciences, data is often kept to support intellectual property claims or new drug application submissions. According to the consultancy The Enterprise Strategy Group, the cumulative amount of archived data in corporations worldwide will grow from about 3,000 petabytes (PB) in 2006 to 27,000 PB in 2010, a nine-fold increase.

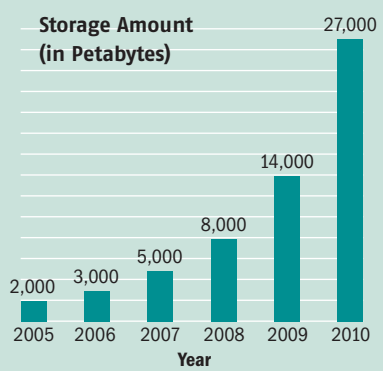
The Impact on Storage Systems is Significant

In the past, most archival data would be moved off of online storage systems and retired to tape and eventually deleted. But today, a large portion of data must remain available online. Additionally, many applications (particularly those that use Web 2.0 and Semantic Web approaches) are designed so the data are available all the time. That means the data are not likely to be taken off primary storage and archived to tape as other data are.

The combination of these factors means life sciences organizations must manage large volumes of data and have the ability to easily add more capacity as demand requires. One way to address the situation is to store data more efficiently. That's an approach being espoused by Vertica, a database start-up co-founded by Andy Palmer, the former CIO and senior vice president at Infinity Pharmaceuticals (See "Pack It In.")

Archiving Takes Off

The need to retain data for longer periods is causing a significant increase in the amount of archived data in corporations.



SOURCE: Enterprise Strategy Group (2006)

EDWARD CARREON

Rico Magsipoc understands the challenge of quickly growing data storage needs. At LONI, his storage requirements have recently skyrocketed.

For many organizations, the capacity planning aspect is a challenge. "When discovery work on a project would begin, we would get space requirements from the scientists," said Peter Herrin, senior systems analyst at Infinity Pharmaceuticals. "They would tell us they would be generating between 20 and 200 GB per month." That's quite a range in capacity.

Infinity Pharmaceuticals selected a storage solution from 3PAR that features a thin provisioning capability. "This allows me to allocate space on a host without using physical storage on the backside," says Herrin. "If [the scientists] only use 20 GB, I haven't shorted myself on physical storage that I can use for other projects."

High Performance with Easier Management

Thin provisioning is just one aspect of easier management. Now, interest is growing in another technology that promises to deliver higher performance storage with simpler management.

For years, storage area networks (SANs) were used only in large enterprises and only for applications that required very high performance. The reason for the exclusivity: SANs were typically much more expensive to buy and, because they used Fibre Channel technology, they were often more expensive to manage since many IT people did not have experience with the technology and thus needed special training.

So the performance SANs deliver is highly desirable. However, many organizations do not want to take on the complexity of using the technology.

The situation is changing. There is now growing interest in IP SANs, which are SANs that use SCSI (Small Computer System Interface) connection technology running over IP. The advantage of IP SANs (also called iSCSI-based SANs) is that they use standard Ethernet infrastructure and low-cost interconnection devices as opposed to traditional SANs that rely on the more complex and expensive Fibre Channel technology.

With IP SANs, the equipment, infrastructure, and management costs are not as high as a traditional SAN. What is especially appealing about the technology today is the introduction of newer products that make use of 10 Gigabit Ethernet (10 GbE) to connect the storage systems. In many cases, this connection speed overcomes bottlenecks that would have required Fibre Channel in the past.

Last spring, Douglas Gibbs, pathology bioinformatics manager at the University of Michigan in Ann Arbor, tested one of these new products from Intransa for use with a medical imaging application. "The 10-GbE infrastructure is what allows the [system] to overcome what would otherwise be front-end 'pipe' barriers to achieving maximum performance," said Gibbs.

With new products like this, the research firm IDC believes the market for iSCSI is poised for rapid growth. In fact,

(CONTINUED ON PAGE 30)

Challenges in Next-Gen Sequencing and Web 2.0 Storage Solutions

Available NOW!

Bio-IT World
OnDemand Webcast

www.bio-itworld.com/webcasts/bluearc

Featured Speakers:

Dr. James Reaney
Director, Life Sciences
BlueArc Corporation

Dr. Scott Kahn
Chief Information Officer
Illumina, Inc.

Dr. Kevin Davies
Editor-in-Chief
Bio-IT World

www.bio-itworld.com/webcasts/bluearc

Next-generation sequencing, and the data that it generates, has presented an incredible data management challenge to the informatics community. Developing a data management plan that addresses both the sheer scale of the data and provides effective management of the data through the informatics pipeline is not easy. While next-generation sequencing holds great promise to make possible a vastly more complete characterization of the human genetic machinery, without a robust storage architecture and efficient data retention strategies few experiments will ultimately succeed. Please join Illumina, Inc. and BlueArc Corporation for a webcast discussion on next-generation sequencing data management strategies.

Participants will learn about:

- The scope of the challenge - Discover how even moderately sized sequencing projects will generate many terabytes of raw and processed sequencing data -- and how larger sequencing efforts will require fundamentally new data retention strategies.
- An analysis of the sequencing informatics pipeline - The speakers will discuss where gains may be made to more effectively manage the data workflow.
- Optimized solutions for informatics architects - See how a high-performance, tiered storage solution can provide future-proof scalability to growing next-generation sequencing research pipelines.

BLUEARC®

illumina®

Pack it in

One way to address the data management issue is to store data more efficiently so that it takes up less space and is easier to query. That is the general idea behind a new database from start-up Vertica.

The company was founded by life sciences veteran Andy Palmer and database veteran Michael Stonebraker. Palmer was most recently CIO and senior vice president at Infinity Pharmaceuticals. He also served as president of the Interoperable Informatics Infrastructure Consortium (I3C). Stonebraker was the main architect of the INGRES relational DBMS, and the object-relational DBMS, POSTGRES.

Most databases are optimized to handle a large number of updates. The Vertica Database is a general-purpose relational database system designed to provide extremely good performance on read-intensive query workloads.

Vertica's Palmer is dedicated to smarter storage.

FILE PHOTO: FURNALD/GRAY

"In many [industries], there are applications and uses of database technology where people spend much more time reading rather than writing to a database," said Palmer. "I figured there was an opportunity to build from scratch an SQL database for read-only mode."

The database organizes data on disk as columns of values from the same attribute, as opposed to storing it as rows of tabular records. This means that when a query needs to access only a few columns of a particular table, only those columns need to be read from disk. Conversely, in a row-oriented database, all values in a table are typically read from disk, which wastes I/O bandwidth.

Storing data in the column-oriented manner improves performance. "Because of the way the data is represented, queries can be completed in reasonable times," said Palmer.

The Vertica Database also uses aggressive compression of data on disk, as well as a query execution engine that is able to keep data compressed while it is operated on. "Because of [the] significant compression, [it] is much more efficient allowing you to keep more data," said Palmer.

According to Vertica, these technologies help execute queries much faster than traditional relational database management systems and require significantly less storage space.

Palmer notes that the technology is well suited to life sciences applications such as those that tag data using the World Wide Web Consortium's Resource Description Framework (RDF). **S.S.**

(CONTINUED FROM PAGE 28)

IDC predicts revenues for these products will have a 73 percent compound annual growth rate from 2006 to 2010.

According to the industry trade group the Storage Networking Industry Association (SNIA), "the factors driving this growth have been the continuing need for IT organizations to do more with less — less capital cost, fewer administrators per Terabyte, less complexity."

Addressing Energy Efficiency

Costs of another type are now starting to become an issue. It takes electricity to run and cool storage systems.

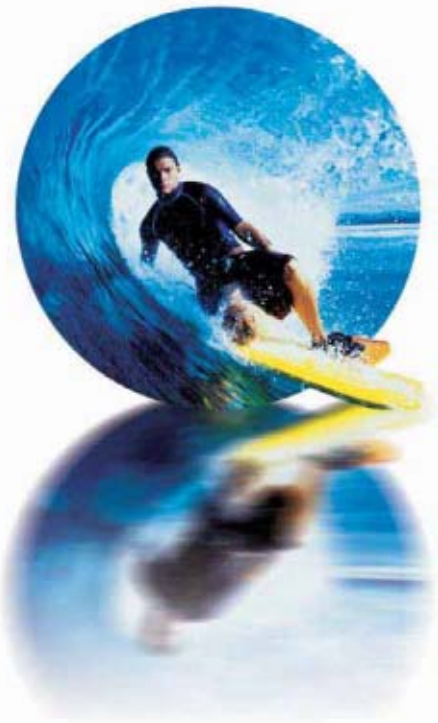
Like most IT and data center equipment, storage devices continue to increase in performance while physically shrinking. While the combination of higher performance and higher densities helps meet the capacity and computational requirements for life sciences research, it also increases the demand for electricity. More power is needed to run the systems and cool the densely packed (and hotter) units.

The emphasis on storage system energy efficiency was made clear in July when EMC announced annual updates to its entire product line. Normally, these announcements focus on improved performance since all of the systems take advantage of the most recent processors and disk drives. This year, performance was certainly mentioned, but so too was energy efficiency.

EMC claimed new Symmetrix, Celerra, and CLARiiON systems reduce power consumption by 33 percent. And a new version of the EMC Centera reduces power and cooling requirements by 67 percent per terabyte.

Many other storage vendors are also targeting energy efficiency. In particular, IBM, HP, EMC, Network Appliance, SGI, Quantum, and Netezza are all members of The Green Grid, a consortium of information technology companies and professionals seeking to lower the overall consumption of power in data centers around the globe.

As data volumes grow, the increased attention to energy efficiency is going hand in hand with simplified management and high performance as key criteria for storage systems to handle life sciences data. ●



Cambridge Healthtech Institute's Seventh Annual

Pep Talk

January 7-11, 2008
Hotel Del Coronado
San Diego, California

The Definitive Protein-Focused Event

Upstream and Downstream Pipeline Breakthroughs

Upstream Discovery • January 7-9

Eighth Annual

Mining the Plasma Proteome
Technical Challenges and Solutions from Discovery-Phase through Development Validation

Eleventh Annual

Protein Expression
Optimizing Outcomes for Maximum Efficiency

Fourth Annual

Peptide & Protein-Based Therapeutics
Pioneering New Frontiers

NEW in 2008!
Downstream Development • January 10-11

Inaugural

Lyophilization
Strategies for Successful Formulation, Cycle Development and Optimization, Regulatory Compliance, Validation, and Scale-Up

Inaugural

Protein Production
Details for Achieving Success

Inaugural

Process Management
Route to Success

Corporate Sponsors:

BlueSky Biotech · CODON DEVICES · GENEART · GenTel BioSciences, Inc. · nematics · nextgensciences · Whatman

Sponsoring Organizations:

CPPC · ProteinSociety.org

Make sure to mention key code U02 when registering!

www.chi-PepTalk.com

Cambridge Healthtech Institute
250 First Avenue, Suite 300
Needham, MA 02494
Phone: 781-972-5400
Fax: 781-972-5425
Toll-free in the U.S. 888-999-6288

Drug Redux

Gene Logic and others recycle, in high tech style.

BY VICKI GLASER

Call it what you will, extracting value from well-studied compounds that bind an interesting drug target, have drug-like characteristics, and offer an attractive safety and pharmacokinetic profile, but, for whatever reason, were not effective in patients, is a lesson in pharmaceutical recycling.

Such terminated chemical entities have failed in clinical studies and are gathering dust on a refrigerator shelf or taking up freezer space. But despite being abandoned, there is intrinsic value in the resources invested in their creation and characterization, the intellectual property (IP) assets they represent, and their potential worth as a source of revenue, whether for out-licensing or as the seed of a future development program or partnering opportunity yet to be conceived.

As pharma productivity lags and truly novel, safe, and effective drugs emerge at a snail's pace, companies are increasingly willing to scavenge through the pharma equivalent of Grandma's attic to search for hidden treasure. What has become commonly known as drug repositioning is sometimes referred to by the more utilitarian term of drug repurposing, or perhaps, reprofiling. Nix the fancy terms, though, and this effort to reposition corporate assets is nothing more than resource recycling, however with a high tech twist.

In 2005, Gene Logic saw the potential for establishing a viable business model around drug repositioning as a niche, value-added activity and built in a creative risk/benefit sharing mechanism. Two years later, with eight pharma partners on board — including Roche, Organon, Pfizer, and recently inked deals with Merck Serono and Solvay — the company views the repurposing of its corporate focus as a successful work in progress. Says Bethany Mancilla, VP of business development and licensing, "We believe the long-term value for Gene Logic lies in drug repositioning." As part of a move toward pharmaceutical development, "We are looking at a number of strategic alternatives that will leverage the value of our genomics business."

Looking back, Marcel van Duin, executive director and head of pharmacology at Organon, attributes the rise of interest in drug repositioning to two main factors: the emergence of enabling technology platforms capable of rapidly evaluating drug properties and demonstrating their efficacy in various models; and the recognition of potential value in compounds that failed for reasons other than toxicity.

Drug repositioning is an opportunity to "drive serendipity," says van Duin. There is a history of successful repositioning — Pfizer's Viagra and Rogaine, for example — that demonstrates that this serendipity exists. Exposing compounds to the full range of technology platforms can drive this process, "but you have to be realistic, it will not dramatically change the numbers of new chemical entities in development."

Early on, Gene Logic applied its core genomics and *in silico* platforms to identify targets for drug screening. These efforts produced the company's toxicogenomics initiative and BioExpress and ASCENTA databases, which have captured transcription profiles of drugs targets in more than 430 disease contexts. During this period, Gene Logic was accumulating a knowledge base of how different compound classes and individual molecules affect therapeutic targets across tissue types. The subsequent acquisition of *in vivo* imaging technology, *in vitro* pathway-based assays, and methodologies for a directed form of metabolomics from Millennium Pharmaceuticals, combined with Gene Logic's bioinformatics capabilities, led to what Mancilla describes as a natural convergence of the companies' business models toward offering a "third route [internal discovery and in-licensing being the first two] for pharma companies to build their late-stage pipelines and improve productivity."

Noel Hall says that identifying novel applications is at the core of the Aspreva business model.

Broadening the Model

Aspreva Pharmaceuticals, from its outset, distinguished itself from other drug repositioning companies by targeting approved or late-stage drugs and searching for value-added alternative indications outside the initial strategic focus. As Michael Hayden, co-founder of Aspreva, told *Bio•IT World* in 2005: 'We won't cost you a cent. We'll take the drug, and we'll share the upside. You focus on the primary market; we'll take care of everything else."

Aspreva acquired the rights to develop Roche's transplantation drug CellCept (mycophenolate mofetil) in all autoimmune disease indications in 2003, and this program has been at the core of Aspreva's growth. However, the company recently announced a reorganization to focus on

(CONTINUED ON PAGE 34)

Drug Redux

(CONTINUED FROM PAGE 32)

core activities that could drive long-term growth, along with a 25% reduction in staff. The announcement followed the June release of preliminary Phase III results showing that CellCept was not superior in efficacy to cyclophosphamide for the treatment of lupus nephritis.

"The philosophy of identifying novel applications is still at the core of Aspreva, but we are changing our tack in terms of how we might access drug candidates," says Noel Hall, co-founder, president, and director. "The company is now also evaluating more traditional approaches to identifying drug candidates," with in-licensing at the forefront of that strategic shift.

In light of Pfizer's history of identifying alternative indications for some of its most successful compounds and the high rate of drug candidate failures at the Phase II clinical trial stage, Donald Frail, head, global indications discovery unit, at Pfizer Global Research and Development, believes that the industry as a whole is exploring the value of alternative indications too late in the lifecycle of a compound.

"There is an opportunity to look much earlier," says Frail, and Pfizer has launched an internal, dedicated research group, "to explore systematically the biology/mechanism of compounds in disease models across multiple indications." Not calling it a drug repositioning initiative, as it includes compounds still in development, Frail prefers "indications discovery."

"I have a bias that drives our strategy," Frail explains. "The most likely identification of an alternative indication will not be through an off-target effect, but rather an on-target effect affecting a biology in another target tissue." One aspect of technology development that could facilitate the search for new indications is in the area of clinical informatics. "A key question, for example, is how to capture physicians observations, whether on marketed compounds or those in clinical trials," — observations of an activity or change in physiologic response that would not be considered an adverse event, and would therefore not require reporting, but that might signal a novel mechanism of action and a beneficial response, says Frail.

Tools for Reuse

Key technology providers have bought into the concept. Melior Discovery applies its theraTRACE indications discovery platform to the systematic analysis of a single compound in multiple in vivo disease models. The company's lead compound MLR-1023 was originally developed to treat ulcers and is now in development for type II diabetes.

> Drug repositioning is a "third route for pharma companies to build their late-stage pipelines and improve productivity."
>
> **Bethany Mancilla**, Gene Logic

Sosei targets its Drug Reprofiling Platforms in two directions — to develop new uses for marketed drugs and drug templates and for compounds that have stalled in Phase II. Several are in Phase II development in new indications including chronic obstructive pulmonary disease, cancer pain, and fibromyalgia.

KineMed leverages its repertoire of *in vivo* pathway-based assays to identify compounds that modulate the kinetics of select metabolic pathways. KineMed has established partnerships in the areas of atherosclerosis, diabetes, osteoarthritis, neurogenesis, and neurodegeneration with companies including Bayer, Merck, Merck Serono, Organon, and Roche.

Gene Logic says publicly that of the more than 70 drug candidates entered into its Phase R drug-repositioning program, it has found a potential new therapeutic use for 25-33% of them. GL1001, acquired from Millennium and repositioned to treat inflammatory bowel disease has completed *in vivo* proof-of-concept testing. Gene Logic is seeking a clinical development partner for the drug.

Thomas Barnes, now senior VP of discovery at Gene Logic, admits that when he was part of the Millennium start-up group, their first presentation of the drug repositioning strategy to venture capitalists was met with skepticism on multiple levels. Could the company identify new indications of value? Could it extract desirable molecules from pharma? Could it structure agreements based on milestones and royalties needed to fuel the business model? And could it convince pharma to assume the risk of taking an abandoned molecule back into development?

"Pharma has learned so well and is accepting the value proposition so much that now we have to compete with their own internal [repositioning] efforts," says Barnes. Outside the United States, Gene Logic found that before it could even try to persuade companies to consider drug repositioning it had to help them "understand the concept that when you kill a molecule it is not necessarily dead."

Two facets of Gene Logic's strategy deserve particular attention. First is its proactive approach. It scours databases for compounds left by the roadside that it deems suitable for its repositioning program. It then approaches a pharmaceutical company with a list of compounds of interest and asks whether they are available.

Second is its risk-sharing business model, in which, "we give the partner the first bite of the apple," as Mancilla describes it. Partners may opt to resurrect a compound and reestablish an internal development program focused on the new target or indication, with Gene Logic sharing in any future success.

"If a partner remains uncertain about a compound's potential once a new indication or therapeutic target has been validated, we may decide to work with the partner to pursue additional testing and further de-risk the program for pharma, explains Stephen Donahue, senior VP of clinical development at Gene Logic. If the partner declines the opportunity, Gene Logic has

the option to move the validated compound forward, in which case the company would likewise share any future gains with its partner.

Corporate Sensitivities

At the heart of a drug repositioning agreement are "drugs that failed in a well-behaved way," says Barnes. They pass through early stage safety trials untarnished, with no significant adverse effects, and either falter in efficacy studies or are discontinued due to reasons unrelated to clinical development, such as budgetary issues, duplicative/parallel programs, or portfolio deprioritization.

Gene Logic asks two things of its partners: provide five grams of material for each compound and reveal the real reason for termination. These discussions may begin to stray into areas of corporate sensitivity regarding intellectual property, strategic planning, and ongoing in-house and collaborative programs.

Organon's van Duin explains that it is not easy to retrieve historical data from discontinued projects. "There is no VP of failed compounds," he says.

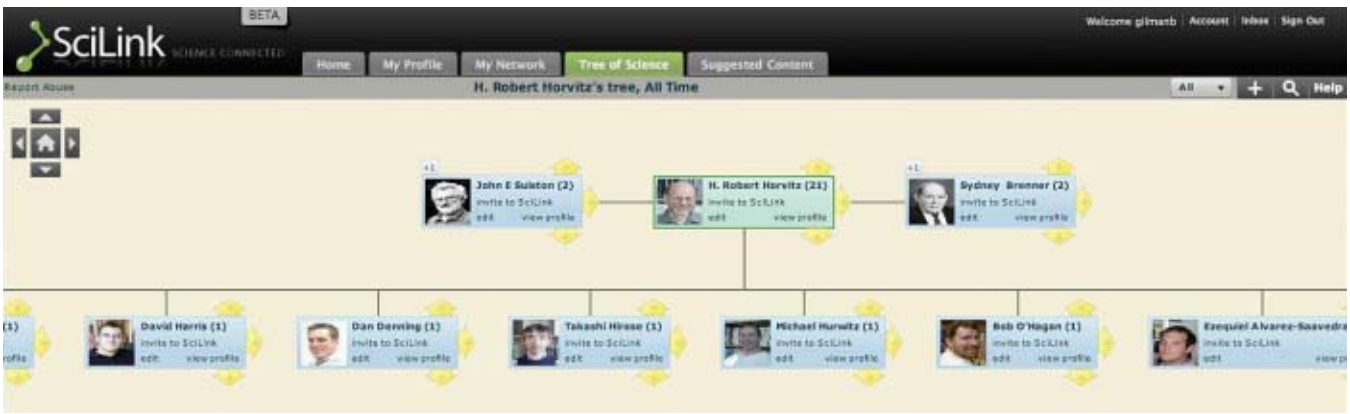
Revisiting and assessing data that were once the focus of intense interest may require information known only by the people who oversaw those projects. Companies need to decide whether to reprise a molecule based on a re-exploration of why the compound was killed in the first place and its predicted or potential new value in relation to the company's current internal development, partnering, and licensing philosophy.

Perhaps one of the most sensitive aspects of collaborative drug repurposing is the act of releasing a pharmaceutical compound to another company for evaluation, especially compounds with a relatively lengthy history of investment, development, and testing. Strategic assets are prized and protected and, as a rule, pharma does not readily share intellectual property, knowledge, or compounds.

But if greater sharing of knowledge and molecules between drug development partners and niche technology providers can bring more new drugs to market and improve productivity, then those barriers may weaken.

Likewise, if drugs that have already received significant investment for proof-of-concept studies, toxicology and pharmacokinetic profiling, and safety testing in humans can be repositioned and successfully commercialized for a different indication or, in concert with efforts toward personalized medicine, for use in a defined patient population, then discontinued and stalled compounds will have the potential to be revenue producers and partnering vehicles.

Regardless of what you call it, more companies are open to the idea of exploring more broadly the potential of proprietary molecules before relegating them to the developmental dustbin. ●



SciLink builds "family trees" for scientists — even Nobel Laureates.

SciLink Scours the Web for Connections

Website reduces the "tedium of networking."

BY KEVIN DAVIES

Brian Gilman has made a lot of friends in the past decade or so, launching successful software companies and playing a key role during the heyday of the Whitehead Institute Genome Center. He hopes his new venture will make him a few more — 7 million more, to be precise.

SciLink — Gilman's latest project — was founded in 2006, and has now grown into "a small child," he says. Upon the company's official web launch in August 2007, Gilman professed to being "giddy with excitement."

The company was bootstrapped off Gilman's previous start-up, Panther Informatics — a 5-year-old bioinformatics consulting firm. Gilman spent four years as a key member of the Whitehead Genome Center, reporting to David Altshuler, director of medical and population genetics and Jill Mesirov, CIO.

"Around 2002-03, lots of high-level consultants were laid off [across the industry]," says Gilman. "I saw that as an opportunity. I was sitting at the Whitehead and needed to make a decision." Gilman didn't really fancy being "a lifer" in academia, so he left, "and started to hire all my friends." Today, Panther has a small full-time staff in Boston plus a dozen or so consultants around the country. "Panther Informatics is interested in a number of different research endeavors: using HapMap data to stratify patients for drug efficacy, tying HapMap data to adverse event reports, and bringing researchers together using advanced web services," says Gilman.

Tree of Science

His latest venture builds on the widespread interest in social networking and Web 2.0 initiatives. But as Gilman points out, "no one was doing professional networking in this space."

SciLink scours the web for publicly available information on the universe of scientists and their publications. "We took all of medline, the .gov and .edu domains, and scraped them to find 5.8 million scientists and their relationships." (Gilman estimates the total universe of scientists is closer to 20 million globally.) SciLink takes those web pages and turns them into resumes, matched with over 18 million research papers.

> "It's a revolutionary way of showing how your career has progressed over time. All this community building is free."
>
> **Brian Gilman**, SciLink

"There's already a professional network that exists inside scientific material," Gilman explains. "So why not take that information and reduce the tedium of networking, which is finding people to connect with?" SciLink allows scientists to "claim" their published material from the database, and easily populate their resume. The site offers a widget — a small piece of HTML script — that can be cut and pasted into the user's own web page to automatically keep an up-to-the-minute CV on their own website. That CV can also be exported into the familiar NIH "biosketch" format. This capability is "very popular among our academic friends," says Gilman, referring to the mandatory biographical standard form that must accompany NIH grant submissions.

In addition, Gilman says, "We're starting to push content like TiVo, where you train our system about what kind of information you're interested in seeing more of." That content includes funding opportunities, jobs, recommended papers, and more. The company has also created private networks for scientific and publishing clients, including Biogen Idec.

Perhaps the most distinguishing fea-

Kelaroo Releases Sequence Profiling Tool

In advance of the release, Kelaroo expands Amylin contract research.

BY KEVIN DAVIES

In advance of its new SeqR product, Kelaroo has expanded its contract research with Amylin Pharmaceuticals, and received an equity investment from the La Jolla, CA, biopharma.

Kelaroo provides cheminformatics and bioinformatics applications and services in support of the drug development work of some 40 biopharma companies, with Amylin being one of the company's first and largest clients. In conjunction with Amylin scientists, Kelaroo programmers have developed a sequence-profiling approach called SeqR, combining machine learning methods and high-performance sequence analysis, suited toward genome-scale data mining and sequence optimization.

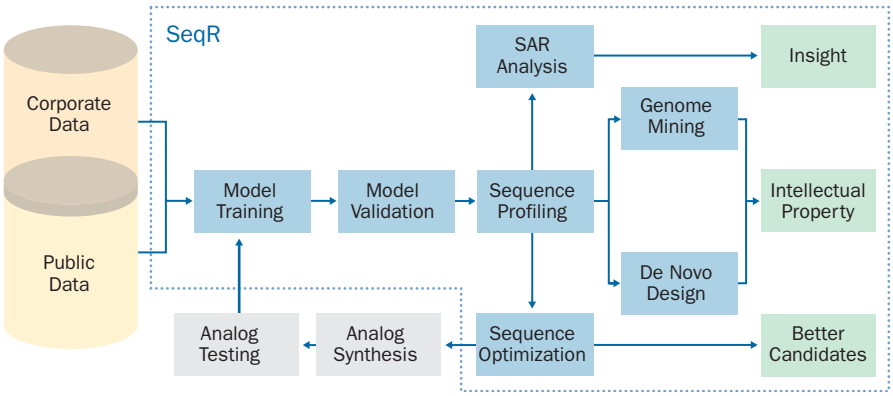
"SeqR is an innovative sequence-profiling approach that combines machine learning methods, automated model validation, and high performance sequence analysis designed for genome-scale data mining and sequence optimization," explains Greg Fond, Kelaroo's director of business development.

"SeqR is not a homology based approach, and thus does not suffer from the limitations of BLAST and other methods that depend upon sequence homology and alignment. SeqR can be trained using public and proprietary sequence activity data, such as that obtained from corporate library screening and physiochemical characterization." In this model training approach, sequence-activity data can be entered into the SeqR machine-learning algorithms to train predictive models. "Training SeqR models on proprietary assay results is a particularly powerful application of the technology," says Fond.

According to Robert Feinstein, cofounder, VP, and CSO of Kelaroo, SeqR has been instrumental in accelerating Amylin's computational hormone hunting and drug development efforts. "Kelaroo's contributions to our research efforts have been substantial and we see the investment as reinforcing a valuable and productive relationship," said Michael Hanley, Amylin's VP discovery research.

Fond says that Kelaroo plans to launch SeqR before the end of 2007. While Amylin remains a strategic partner in the development and validation of the SeqR technology, Fond says that Kelaroo will leverage its existing customer relationships during the product's initial launch, and will be offering incentives for early adopters of this technology.

The SeqR technology was developed "to provide a revolutionary platform for mining of genomes and proteomes for sequences of interest," according to Andrew Reum, Kelaroo's president and CEO. "I believe the promise of bioinformatics is embodied in the utility of this system. SeqR has tremendous potential for any company working on peptides, proteins, or biologicals. It is particularly satisfying that Amylin, as one of our most loyal customers, will be a key beneficiary." ●



SeqR mines genomic and proteomic data to train predictive models for drug discovery.

ture of SciLink is the Tree of Science. Says Gilman: "It's a family tree of science — a way to visualize relationships and careers over time." The Tree of Science is free, and allows SciLink members to easily build a graphical, historical resume illustrating links to colleagues throughout their career along a timeline. In the first week since its launch, Gilman says more than 600 people built their trees adding at least three connections. As users scroll through their timeline on the X-axis, the Tree of Science presents a movie, illustrating collaborations, mentor-student and peer-to-peer relationships. "It's a revolutionary way of showing how your career has progressed over time," says Gilman. "All this community building is free," he says, although SciLink will charge for integrating the movies into other documents.

Gilman enjoys good relationships with other commercial groups exploring scientific social networks, even if he doesn't necessarily agree with their strategies. He is currently working with academic and commercial groups to form a "Science 2.0" publishers group in Boston to share ideas, techniques, and foster collaborations.

With biopharma constantly looking for experts and consultants, Gilman says "We're spending money to find experts and figure out what they're good at." Gilman believes SciLink will be a powerful recruiting and collaboration tool — and that those services, along with advanced subscription registrations and advertising income, will connect the company with success. ●

Life on EDC Street

Analysts see Phase Forward, Medidata leading the crop.

BY DEBORAH BORFITZ

With adoption of electronic data capture (EDC) at an all-time high, the question now is which EDC vendors will dominate the market, and how soon? "In terms of data acquisition, management, and analysis," says Curt Sprouse, president of Boston Market Strategies, EDC is superior to the old methods. But, "The future of EDC needs to address sites' needs and greater compatibility with investigators' workflow," says Chip Kalfaian, director of the life science solutions division of Cranford, NJ-based Paragon Solutions.

Analysts contacted for this story generally consider Phase Forward to be the market leader, largely due to its stability and familiarity. "Big pharma trends toward the security of the biggest players, especially Phase Forward," says Kalfaian. "The result is that smaller EDC players are forced to build their businesses one trial at a time." And that is both costly and difficult for small companies to bear.

Publicly traded Phase Forward is a known entity, agrees Sprouse. It wins business "by default" with large companies that aren't apt to swap vendors "unless something is wrong or broken." And Phase Forward has repeatedly proven its capabilities. "Those attributes are important to big pharma. The EDC space is scary to them...they're concerned [other EDC] companies won't be in existence in three years."

However, Chris Connor, senior research analyst with Framingham, MA-based Health Industry Insights, notes that Medidata has made a "huge impact" in a market seeking scalability, and Sprouse agrees. Medidata has done well breaking into the top tier, Sprouse says. "It has won over some big accounts" and is positioned to gain business "if it can more effectively address installation, services, and support."

Scalability doesn't necessarily put smaller EDC companies at a disadvantage, Connor says. Phoenix Data Systems (PDS), for instance, enjoys success by offering an appealingly easy-to-use, affordable product for Phase I trials. "PDS selectively goes after high-value studies, helping to establish itself as a dominant player in the small, Phase I market. It gets follow-on studies — Phase II, III, and IV — using the same technology," says Connor.

Kalfaian, though, considers Oracle the number three vendor behind "safe and trustworthy" Phase Forward and "near second" Medidata, even if its EDC credentials remain a bit cloudy: "You can't ignore the pharma giants who continue to work with Oracle — the same giants that helped establish Oracle Clinical in the back-end data management space. Oracle's scheduled year-end RDC release, if successful, will make it the [EDC] product of choice for many sponsors with a heavy investment in Oracle Clinical."

Radar Love

The analysts agree that both strengths and weaknesses of different companies can be traced to their public or private status. "All the publicly traded companies have the advantage [when big pharma looks for a track record,] because of public disclosure requirements," says Sprouse.

Yet OmniComm is the only other public company currently on the radar for Sprouse, who is looking on behalf of iBall (a Boston-based image management platform) for partnering opportunities with top-tier EDC vendors. "It has the most recent technology platform and does a good job with integration and scale" — two areas that Sprouse's clients have found to be challenging and sometimes problematic for Phase Forward and Medidata. "The functionality, usability, and integration features of [OmniComm] technology are the best in the market today," he says.

Connor, though, says being a publicly traded entity can be a hindrance because of the investment required to take advantage of exponential growth in the market. "The way the EDC market is structured, it's pay-for-performance. Companies don't recognize a return until the study goes live...and investors don't understand that."

OmniComm is doing "fantastically well" and grabbing attention, Connor acknowledges, but argues that PDS may be doing better because it's privately held. Public ownership puts the squeeze on profits and operations. "There are lots of opportunities, but often no way to grab them."

Misfortunate Events

Several publicly traded EDC companies have felt the squeeze that Conner mentions. Datatrak recently laid off 17 employees and in its last quarter reported a $3 million loss on revenues of $3.1 million. At etrials, sales declined 31% following the departure of co-founder/CEO John Cline.

ClinPhone profits also fell below expectations, due largely to an interactive voice response server glitch that did not affect the EDC business with its own servers, says David Stein, VP data management solutions in ClinPhone's product management division. But an increase in sales of EDC software service rather than license, and the weak dollar also hurt business, he adds.

The leading EDC players are not exempt, either. Phase Forward and Medidata both address scaling and integration with a costly, labor-intensive approach, says Sprouse. "[The price problem] can be fixed and, to retain its number one position, Phase Forward will have to fix it." But transitioning to a new platform in a growing market is easier said than done, Sprouse says. Kalfaian says that Medidata has been hampered by service and performance issues, perhaps due to its swift ramp up.

Datatrak stands out in the customer service arena, "assigning knowledgeable study managers to their clients" much as central laboratories do with project managers, says Kalfaian. But the company has been hurt by its "conservative approach to technology," prompting it to adopt an entirely new platform.

Datatrak's acquisition of ClickFind last year was laudable, giving the company "more scalable" EDC technology, says Connor, "but the market is more about execution and delivery than about software functionality." The company forced customers to make a mass migration to the new technology platform, essentially

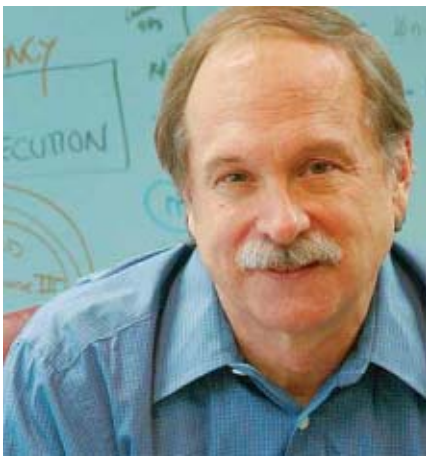
Decisions, Decisions on Adaptive Research

Trial metrics are key, says Health Decisions' CEO.

BY DEBORAH BORFITZ

Adaptive trials require fast decision-making by study sponsors and research teams, and Michael Rosenberg built a clinical research organization 18 years ago to help. The "tactical component" of adaptive trials — continuously tracking enrollment and better matching monitors to need — isn't a new idea, says the founder and CEO of Health Decisions (Chapel Hill, N.C., and Oxford, U.K.). But, like sample size reassessment and dose finding on the strategic side, it remains the approach less taken because of hopes placed on Web-based electronic data capture (EDC).

Fast, clean data and performance metrics are the key criteria for doing adaptive trials of any type, says Rosenberg. That also describes the two major disadvantages of EDC. "Someone needs to enter data by hand, and that inevitably leads to delays. The interval between when data is generated and when it's entered can take a couple of days and, more often than not, a couple of weeks," he says. Moreover, Rosenberg continues, "They miss the most important management component of a study, which has the most profound implications for study timelines and budgets." Years ago, Health Decisions shaved 1.6 years off a five-year timeline and $32 million off a $100 million budget for a six-nation Alzheimer's study using performance metrics to guide managerial decisions.



Rosenberg values the tactical component.

That was even before better data input devices, such as Health Decisions' optical SmartPen, came along, says Rosenberg. The SmartPen records strokes as case report forms are filled out and then docks in a computer station, transmitting data from the device to a central, Web-based database within seconds.

The combination of strategic and tactical components of the adaptive approach is "no less important than the introduction of the assembly line was for the auto industry," says Rosenberg. All trials today should incorporate some elements of adaptive trials, he says, "especially on the tactical side. If you don't consider these elements, you're leaving something on the table."

The inefficiencies of conventional study management techniques are mind-boggling, says Rosenberg. It generally takes a week to access key study progress metrics like site enrollment statistics. "We keep track of what's going on every single minute at every single site and make this [information] available in real time to all study personnel, 24/7," says Rosenberg.

The upfront, per-patient cost of the adaptive approach is no more expensive than traditional methods, says Rosenberg. The savings are on the back end. The hard part is the number of unknowns, such as the duration of a trial. "Sponsors sometimes get overly optimistic about how long it will take to recruit patients," says Rosenberg. "The risk is great so the payoff is great, too. To minimize risk, we make decision-making the best it can be." ●

---

re-opening the RFP process while eliminating Datatrak's incumbency status.

Tiered Identities

Recovery from these setbacks is a driving force in what some see as the emerging hierarchy of the EDC market. Some smaller companies, like Nextrials, could be counted among "second tier" organizations with solid technology platforms, says Sprouse. But he is unsure whether either Datatrak and ClinPhone are positioned "to be a dominant player in the market."

ClinPhone, unsurprisingly, disagrees. Bidding on projects is usually against Phase Forward, Medidata, and Oracle, says Stein. The company last year acquired DataLabs, considered a strong second-tier player by Forrester Research. The vision moving forward includes continuing to focus on the user experience, driving toward "flawless delivery" on its growing ASP business, developing a next-generation design tool, and "advancing the integration capabilities across ClinPhone's eClinical product portfolio."

Etrials also remains optimistic. The company reports that it has a new CEO, COO, established a client services division, and doubled the size of its sales force.

Connor thinks the custom of lumping EDC vendors into "tiers" does the market a disservice. "One size does not fit all," he says. "There are plenty of cases where, if you matched up Phase Forward and PDS, Phase Forward would unquestionably be the way to go. [But] if pricing is an issue...PDS may be the only way to go."

In terms of total market share, Connor gives top billing to Phase Forward and Medidata, followed by etrials, Datatrak, and companies like Fremont, CA-based Velos that specialize in investigator-initiated trials. "Velos is the most successful company no one has ever heard of. It has tens of thousands of studies."

Although consolidation is likely, including acquisitions by contract review organizations (CROs), smaller EDC vendors like Target Health (New York, NY) continue to pop up in the market with dynamic EDC tools, says Kalfaian. "Smaller companies don't need that many trials to be viable. They can book a couple of CROs and biotechs and be good for a while. It's hard to shift technologies once a trial is underway." ●

New Products

Linkage Analysis Help

The Human Linkage-12 Genotyping BeadChip is Illumina's first standard panel to take advantage of a twelve-sample format for linkage analysis. The HumanLinkage-12 BeadChip offers the lowest cost per sample for linkage analysis; PCR-free analysis increasing accuracy, ease, and speed; uniform marker distribution; and excellent SNP content.

Product: Human Linkage-12 BeadChip
Company: Illumina
Available: Now
For More Information: www.illumina.com

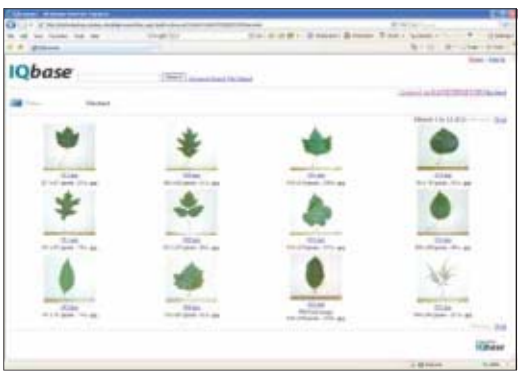
Data Mining and Visualization

The AKS v.2.1 knowledge mining system is now available as a subscription service. AKS (AlmaKnowledgeServer) v 2.1 from Bioalma allows users to uncover bioentity relationships that exist in over 8 million PubMed abstracts, and now includes chemical information for millions of compounds and links to external databases such as PubChem. Version 2.1 also boasts query alert emails, new search and export options, and new trend analyses, all available though an internet subscription.

Product: AKS version 2.1
Company: Active Motif
Available: Now
For More Information: www.activemotif.com

Image Management for Large Volume Collections

IQbase 2.5, a scientific image management software, enables organizations to store, query, and share large numbers if images and related data. IQbase now supports new file formats including JPEG 200, Zeiss LSM, Leica LEI, Image-Pro SEQ, and Metamorph SDK in addition to other standard formats. Users can search stored images by keyword or via context-sensitive search parameters, and images can be shared using PDF, PowerPoint, and Web-based sharing tools. Images can also be archived via the internet using online upload capabilities.



Product: IQbase version 2.5
Company: Media Cybernetics
Available: Now
For More Information: www.mediacy.com

Advertiser Index

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

The Russell Transcript

Optimata: Eyes on the Prize

JOHN RUSSELL

Remember the joke about the fellow who walks into a psychiatrist office and complains that his head hurts from banging it against a wall? And the psychiatrist says...

Optimata, a seven-year-old predictive modeling specialist, hasn't so much given up banging against the fee-for-services wall (it will still do some of that), but it has decided to shift focus to a potentially more lucrative wall — drug repurposing and development. It's likely a smart move and a good bet that other systems biology firms will follow suit, finding their heads equally sore.

Founded in 2000 by pioneering biomathematician, Zvia Agur, Israel-based Optimata has spent the intervening years developing its Virtual Patient platform, a powerful predictive engine perhaps best described by Agur: "In cancer we describe the whole disease dynamics. So when we're talking about the growth of very miniscule tumors, they have geometric arrangements, cell cycles, diffusion of nutrients, and the drug eventually in the tumor. We create a detailed mathematical description of all that. Just to give you an idea, to translate only angiogenesis into biomathematical algorithms is a question of about 2.5 man-years work." (See "Optimizing Optimata," *Bio•IT World*, November 2005, for more background).

Input key patient data and Virtual Patient is able to run virtual trials, testing various drug regimes. Last October, Optimata reported results of a proof-of-concept project with cancer researchers at the Nottingham City Hospital, U.K. The Virtual Patient platform was able to predict breast cancer patients' response to chemotherapy drugs with 70 percent accuracy. That's substantially better than what most oncologist achieve, says Optimata. Another pilot with Lilly, this time supporting development of a new compound, went well and Lilly extended its collaboration last January.

Turning stellar science into cash, however, has proven difficult.

Finding the Right Business Model

In late 2005, Guy Malchi was brought in as CEO, "to find a business model that will work. That's the reason I joined. It took a long time to analyze what's going on out there in terms of how to create value. The first thing we wanted to do is to engage with companies that would benefit from the technology in order to make it not a theoretical exercise but to understand exactly the needs of the other side of our partners, and also understand our capabilities versus those needs," he says.

Malchi joined Optimata from the London-based European Life Science practice for TEFEN Ltd., a global management consultancy firm. He is a founding partner of the practice and for seven years "substantially grew" the business via collaborations with global pharmaceutical companies such as Johnson & Johnson, Pfizer, AstraZeneca, Schering-Plough, and Yamanouchi.

Now, says Malchi, "I believe we have identified what we hope is the sweet spot, which means to try to identify these continued oncology compounds that were proven to be safe in man, so passed successfully Phase I [but failed for some other reason]. We would then license them, and use the Virtual Patient technology to find a new direction in terms of indication, patient population, drug regimen, and then actually go out to carry out the Phase 2 that's to validate our prediction. We will be able then to out-license the compound."

> "We had to change the DNA of the company from a company which is science-based, science-oriented, to a science-based marketer-oriented organization."
>
> **Guy Malchi**, Optimata

Malchi says Optimata has a list of "less than a hundred compounds that we believe that we can re-purpose, and we are now in the process of approaching those companies." To drive that effort, Optimata just hired a VP of product development (Alex Chausovsky).

Certainly big challenges remain. Drug repositioning is hot, and many are dabbling in it (See p. 32). It's not yet clear how many are succeeding.

"I think the biggest challenge is to continue to enhance our core competency, whilst introducing completely new capabilities to carry out Phase II [activities]. Israel is very good in drug discovery, but we don't have a lot of clinical development expertise. It means that the pool of knowledge in Israel is limited and we have to look outside," says Malchi, who notes additional funding will also be needed to carry out phase two trials.

Stay tuned. (A more complete interview with Optimata's CEO is available at www.bio-itworld.com)

The Lasergene 7.2 suite of sequence analysis tools shines new light on the order of nature.

Lasergene has long been a workhorse meeting molecular biologists' needs for desktop sequence analysis. Whether for Windows or Macintosh platforms, discover how Lasergene software can assist in your sequence analyses and help you discover the order of nature.

Contact us for a free 30-day fully functional trial version of Lasergene v7.2

- Handle large bacterial genome-sized projects faster than ever
- Assemble 454 Life Sciences™ and/or Sanger runs generating Flowgrams and trace files
- Sequence assembly and alignment
- SNP discovery and detection
- Automated virtual cloning
- Protein structure prediction and gene discovery
- Primer design
- Available for Windows or Macintosh
- Microsoft Vista™ compatible

For more information, or to order today, call us toll-free in the **US and Canada at 1.866.511.5090,** or in the **UK at 0.808.234.1643**

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

DNASTAR Lasergene™
www.dnastar.com/lasergene

WE CAN. Yoh has been delivering high-impact talent to America's leading companies for the past seven decades – longer than any of our competitors. These customers count on our experience, time-honored processes, responsiveness and market intelligence. And our delivery of talent that is 24.7% better than the competition*.

Focus on your job. And your future.

YohClinical
Know greater talent.®
yohclinical.com