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

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

We hope the magazine is meeting your expectations. We encourage you to share this issue with your team to stimulate further conversations. Share this issue.

We are genuinely excited about the prospects for this field —and the magazine —in the coming year. We've cemented our position as the flagship publication of Cambridge Healthtech Institute (CHI), and will continue to play a prominent role in hosting and reporting on the best CHI conferences.

We pride ourselves on publishing critical insights and analysis of innovations across the drug discovery pipeline— from molecular modeling of popular drug targets to biomarkers that discriminate cancer responders, and from data handling for next-generation sequencing to new strategies for increasing the speed and efficiency of clinical trials. We will work hard to surpass those stories, continuing our pursuit of the most critical tools and strategies that epitomize the world of "predictive biology."

We hope to continue to engage you with our editorial content, and within our network, and as always, we welcome any and all comments or suggestions editor@healthtech.com

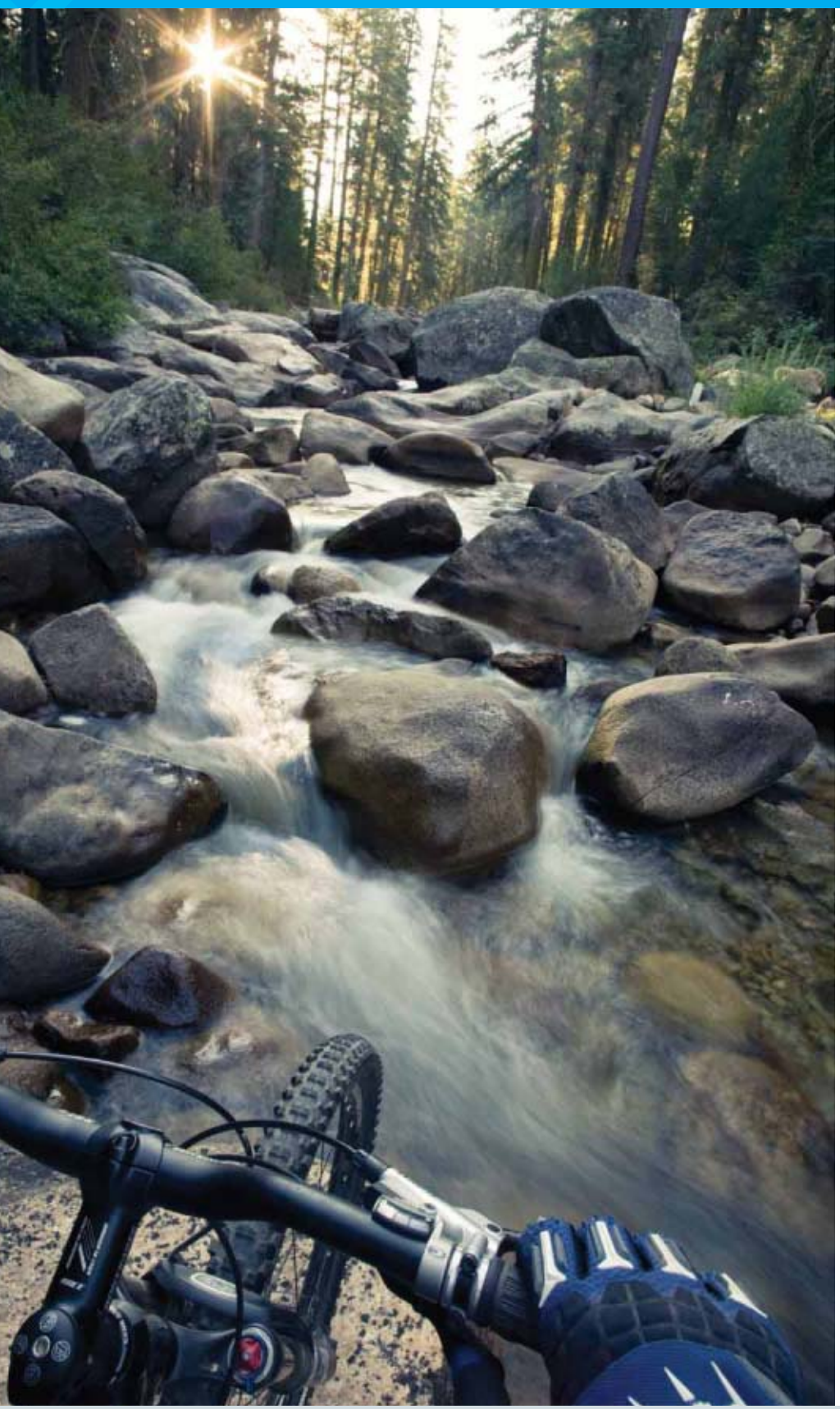
Bio-IT World reports the following digital circulation figures as verified by the Qmags Self-Audit procedures:

November/December 2008 Issue	
Total Qualified Digital Subscribers:	8,544
Biopharma/Chemicals	2,021
Biotech/Bioinformatics	2,176
Acad/Gov, Clinical Trial, CRO	2,742
Hosp Lab/Healthcare, Other	1,604
October 2008 Issue	
Total Emails Sent:	7,700
Undeliverable :	99
Total Delivered:	7,584
Removed:	33

NAVIGATION KEY

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

Go beyond the everyday. EVERY DAY.



Symyx Notebook

The Freedom to Experiment.

Whether it's on the trail or in the lab, you want the freedom to take new approaches, routes, and paths to your goals. That's why there's Symyx Notebook. It's the only electronic laboratory notebook that can be deployed across the enterprise in multiple scientific disciplines. With Symyx Notebook, research teams share a single application to document, work, collaborate, and speed the experimentation workflow.

Symyx Notebook streamlines the capture of all experimental information and intellectual ideas. Everyday tasks such as data capture and note taking are optimized and automated. All of which gives you the time and freedom you need to experiment—and get back to doing science.

To learn more, visit www.symyx.com/notebook6

Symyx®

Cambridge Healthtech Media Group

www.bio-itworld.com

Bio·IT World

Indispensable Technologies Driving Discovery, Development, and Clinical Trials

NOV. | DEC. 2008 • VOL. 7, NO. 9

Secrets of Your Sequence

A first person, first anniversary appraisal of personal genomics 16

> SUBJECT: DAVID N. ADDISON
> Ancestry: CEU
> Disease: Alzheimer's
> Gene: APOE
> Chr: 19q13.2 (50103781)
> SNP: rs429358 / rs7412
> Genotype: CC/CC (E4/E4)
> OR: 11.4-15.5

WYETH'S DISCOVERY INFORMATICS SOLUTION 22

EUROPE'S ECONOMIC BIO-COMMUNITY 28

IS COMPLETE GENOMICS A GAME CHANGER? 8

It's about integration.

Which is why it's about time you invested in a Thermo Scientific Laboratory Information Management System (LIMS). The most successful businesses are those that utilize technology to bring the full scope of the lab into the business – by integrating people, data systems and instruments. We build the integration capabilities you need into the core of every Thermo Scientific LIMS so you can make faster and better informed business decisions, decrease your time to market and improve your margins.

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

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

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

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

Moving science forward

Thermo SCIENTIFIC

Part of Thermo Fisher Scientific

Contents [11-12·08]

Cover Story

Keeping Score of Your Sequence

Twelve months after the debut of the first personal genomics services, Kevin Davies expectorates and explores the pros, cons, and controversies of "23 *et al.*" 16

Up Front

Computational Biology

Computational Development

IT/Workflow

Clinical Research

Economic Development

In Every Issue

Cover photo by Alfred Pasieka / Photo Researchers, Inc.

First Base

Decoding a Mystery

KEVIN DAVIES

It is difficult to watch a company with the innovation and imagination of Iceland's deCODE Genetics struggle to sustain its business given its role at the vanguard of genomic research and medicine. However, last month deCODE retained the Stanford Group to conduct a strategic review of its businesses, which encompass gene discovery, drug development, and molecular diagnostics, in hopes of charting a course to short-term survival and long-term success.

In recent years, but especially since the 2007 boon in identifying gene associations for common diseases, CEO Kari Stefansson and colleagues have contributed more robust discoveries than any other organization, which has spurred the launch of a series of diagnostic kits. But if this decade has taught us anything, it is that translating gene discoveries into drugs is no sure thing. deCODE has managed to push three promising cardiovascular drugs into the clinic, but the most expensive parts of those trials lie ahead. With depleted funds, the current economic mess—Stefansson faults deCODE's American investments more than his own country's plight—brings additional problems.

FILE PHOTO

A younger Kari Stefansson

In a typically candid admission to me shortly before the extent of Iceland's financial woes became known, Stefansson acknowledged the mounting business pressures facing his company. "You can argue that our company was founded about six years too early, because the technology to systematically isolate disease genes wasn't available really until a long time [later]," he said.

"We've performed very badly in telling out story in the investment community," he continued. "When you've been running a company like this for a considerable number of years before the technology is able to deliver, you end up collecting legacy, you collect debt. We have a convertible bond debt of $225 million, which is basically the cost of running the company..., I'm not happy with it, but am I deeply concerned? No. The sales of our diagnostics are picking up. The company has a very realistic business plan that shows we're going to become profitable in a very short period of time. The markets and the rest of the world will appreciate it pretty soon."

Although Stefansson asserted that his company was "leading the field of human genetics," he said unnamed critics were "trying to undermine our credibility," for example by "insisting that our diagnostic tests are not measuring any real risk." The company's latest test for breast cancer is no exception—few if any of the seven markers in that panel offer a biological rationale for assessing cancer risk. But deCODE insists that its markers are the fruit of replicated studies on some 100,000 women and are capable of identifying several-fold increases in a woman's cancer.

November marks the first anniversary of the launch of the personal genomics service deCODEme. Stefansson said his original business proposal "was always to use genetics for preventative health care" (see "Keeping Score of your Sequence," page 16). Stefansson believes that whole-genome sequencing will replace genotyping within the next 2-5 years. He is proud of a new deCODE method to calculate long haplotypes in the Icelandic genome to aid in the discovery of rare sequence variants. "We need to sequence about 2,000 Icelanders to be able to impute the sequence of the entire nation, because we have the genealogy of thousands of people," he said.

We've had some interesting jousts with Stefansson, who keynoted the 2006 Bio-IT World Expo, over the years. Hopefully, deCODE's reorganization, scheduled to be announced in November, will do the trick.

Bio•IT World's New Recruit

Alert readers will notice a new byline appearing in this issue. We are delighted to welcome Alissa Poh as the latest member of *Bio•IT World*'s editorial team. Alissa will be editing PharmaWeek, one of our most avidly read eNewsletters. Alissa hails from Malaysia and studied biochemistry in the U.K. before moving to the U.S., continuing her studies at Dartmouth College and finally the prestigious University of California Santa Cruz science writing program. She has penned some compelling stories around pharma strategy and innovation, two of which are included in this issue (see pages 15, 22) . You'll be seeing her byline much more in these pages and online in the months to come.

Even more alert readers will note that this issue takes us through to the end of 2008. In 2009, we will publish bimonthly issues of *Bio•IT World* while expanding our ongoing coverage of the bio-IT industry through a range of web and online media. We look forward to seeing you in a prosperous New Year!

Company Index

CORRECTION: In the introduction to the October 2008 cover feature "Eric Schadt's Integrative Approach to Predictive Biology" (page 21), the reference to Rosetta Genomics should have read Rosetta Inpharmatics.

Advertiser Index

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

VOLUME 7, NO. 9

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

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

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

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

CAMBRIDGE HEALTHTECH MEDIA GROUP

Bio-IT World®

Indispensable Technologies Driving Discovery, Development, and Clinical Trials

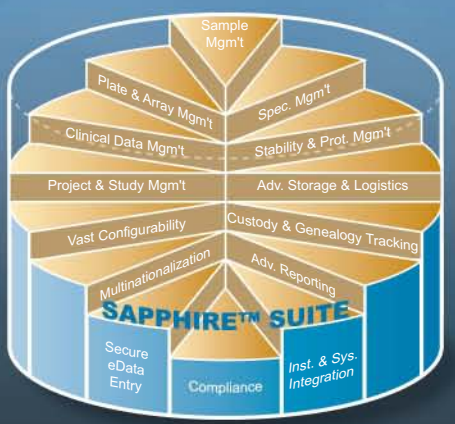
DISCOVERY · DEVELOPMENT · BIOBANKING · FORMULATION · QUALITY MANAGEMENT

LABVANTAGE UNVEILS SAPPHIRE™ R5

A WORLD OF LABORATORY INFORMATION MANAGEMENT OUT OF THE BOX

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

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



Configure your success with the Sapphire™ Suite at:

www.labvantage.com/suite

Keyword: OutOfBoxLIMS

IBM Business Partner

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

LabVantage

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

Complete Genomics Targets 'The First $1000 Genome'

Sequencing service is on track for Spring 2009.

BY KEVIN DAVIES

Complete Genomics emerged from stealth mode in early October brandishing an audacious service model for wholesale next-generation sequencing, with its first human genome already assembled and the CEO's pledge to reach the magical "$1000 genome" price point as early as spring 2009.

"Our mission is to be the global leader in complete human genome sequencing," chairman, president, and CEO Clifford Reid told *Bio•IT World*. "We are setting out to completely change the economics of genome sequencing so that we can do diagnostic quality human genome sequencing at a medically affordable price. Essentially, [we'll] transition this genome sequencing world from a scientific and academic endeavor into a pharmaceutical and medical endeavor."

Based in Mountain View, Calif., Complete Genomics has raised $46 million in three rounds of financing since its incorporation in 2006. Complete Genomics will not be selling individual instruments, but rather offer a service aimed initially at big pharma and major genome institutes. The company is building what Reid calls "the world's largest complete human genome sequencing center so we can sequence thousands of complete human genomes, so that researchers can conduct clinical trial-sized studies."

If all goes according to plan, that 32,000-square-feet, $75 million facility will deliver 1,000 human genomes in 2009 and an eye-popping 20,000 genomes in 2010. But that's just the beginning. The firm plans to build ten genomes centers in the U.S. and abroad over the next five years in partnership with various organizations and foreign governments.

The $4000 Genome

Although the data are unpublished—advisory board member Leroy Hood admitted he hadn't seen the genome assembly results at the time of launch—Reid says his team produced its first human genome sequence last July. "It's getting a bit long in the tooth already," he jokes. "The total materials cost was about $4000." The genome coverage was 22-fold from a total of 67 gigabases (Gb) of mapped reads. "The speed of the instrument is about ten times as fast as ABI and Illumina," Reid claims. "This [project] ran four instruments for one run of a week. This is a 28- instrument-day experiment. By the launch of our product in Q2 [of 2009], it will be a 4-instrument-day experiment.

When the product is launched next spring, "we fully anticipate the materials cost of that genome will be just under $1000," Reid says. "We're going to price the genomes at $5000 each, which covers of course not only material but also instruments and labor and overhead." Reid admits it's an incomplete measure of cost, but it has become the industry's standard accounting method. By those criteria, "It will be the first $1000 genome," says Reid.

The Complete Genomics technology is based on co-founder Radoje Drmanac's work in sequencing-by-hybridization using a ligation strategy and gridded arrays of up to one billion DNA "nanoballs." Reid's expertise in computer science and Drmanac's in biochemistry proved to be a perfect complement. "The convergence of biotechnology and computing has really enabled a whole new generation of DNA sequencing that's going to change the world," he says.

First Service

Complete Genomics is not the first company to explore a service model for genome sequencing (see, "Genome Corp. Born Again," *Bio•IT World*, Jan. 2008), but it is the first using a next-gen sequencing technology.

Reid is setting out to completely change the economics of genome sequencing.

Reid cites two main reasons for choosing a services model. First, he sees the key market as large pharma conducting clinical trials. "The pharmaceutical market has declared very clearly they don't want to buy instruments," says Reid. "They want to buy services, so that they get the data that enables them to do the discovery and development work, rather than have to own and operate a large-scale genome sequencing center."

A second consideration is that the new sequencing technologies "generate a breathtaking amount of data," says Reid. "Simply selling 10 or 20 instruments to a company doesn't solve the problem. You then have to be able to mange huge volumes of data. We are putting in a Google-style data center to manage the data." (See, "Petabytes of Data")

Reid plans to build a further ten centers—for about $50 million apiece—in the U.S. and abroad over the next

(CONTINUED ON PAGE 10)

Need to assess the novelty of substances?
SciFinder is the answer.

It includes CAS REGISTRY,[SM] the most comprehensive substance information available, integrated with relevant journal articles and patents.

Give your research team the highest quality and most timely scientific information resource.

Make SciFinder an essential part of your research process.

For more information about SciFinder, visit www.cas.org or e-mail help@cas.org.

an essential

SciFinder®—Part of the process.™

CAS | SciFinder®

www.cas.org

CAS is a division of the American Chemical Society

(CONTINUED FROM PAGE 8)

five years, in partnership with other companies, research organizations, and countries. Those ten genome centers will produce about one million genomes over the next five years. "A nice way to think about 1 million genomes is 1,000 people with each of 1,000 diseases. By the time we've done that, we will understand the genetic basis of all the important human diseases," he says.

Scale Up

The near-term goal for 2009 is to focus on pilot sequencing projects for the commercial and academic communities to validate the technology and establish workflows that will form the operational blueprint for expansion. Ten percent of the firm's sequencing capacity in the next two years will be devoted to a collaboration with advisory board member Hood. The Institute of Systems Biology president is a partner with the Government of Luxembourg on a $200-million biobank and personalized medicine project.

As for targeting pharmaceutical companies, Reid predicts two key groups of early adopters—companies pursuing cancer and mental illness. Both groups of diseases have a strong genetic component. Says Reid: "To date, the industry has not been able to find the rare variants that are causes of diseases and drug response. That's a new capability we're bringing."

Another enticing constituency for Complete Genomics is the personal genomics or consumer genomics market. Reid agrees: "Knome and 23andMe and Navigenics and all those guys will essentially buy genome services from us and add a lot of value [and] transfer it on to the consumer population." •

Editor's Note: Complete Genomics CEO/Chairman Clifford Reid will be a keynote speaker at the 2009 Bio-IT World Expo (April 27-29, 2009).

Petabytes of Data

The Complete Genomics platform hinges on exquisite precision in manufacturing and arraying "nanoballs" of DNA. But it will be critical to manage gargantuan quantities of data. The task of building the data center falls to vice president of software Bruce Martin, a former executive with Sun and Openwave.

"I've built a team that is a little microcosm of what you see in the rest of the company," says Martin, including bioinformaticians who worked with Craig Venter on genome assembly and the HapMap project, as well as experts in data mining, indexing databases, and high-throughput computing.

The imaging steps involve measuring hundreds of millions of spots. "We are currently generating close to a gigabit a second off the imager, and that's going to go up by a substantial amount in the next year," says Martin. "I have not only an extremely interesting computational challenge here, but there's just a bandwidth problem... You can't store images at that rate onto disk drives without spending a king's ransom in storage."

Martin says his group has had "a very successful run" with a clustered storage system from Isilon, which he likes for its "very high performance" and ability to scale to multi-petabyte file systems. "You can manage it with a very small footprint of staff. The Broad recently deployed them as well. I couldn't say who got there first. We both basically have selected them for similar reasons."

Due to space, power, and cooling considerations, Martin is exploring options with several high-density blade vendors. "We want to pack as many cores and as much memory into as small a footprint as we can for economic reason," he says.

Martin says he's made "a significant investment in an aligner" for rapid genome alignments that can scale to thousands of processors. "I went out and found some very significant expertise in Silicon Valley in terms of high-speed, large-scale search and indexing. We have many of the leading companies in the world in that area."

If the ramp up for 2009 sounds daunting—1,000 genomes in a center housing 5 petabytes of data—the specs for sequencing 20,000 genomes in 2010 are positively frightening. "We'll probably be in the 60,000-processor and 30-petabyte range in that time frame," says Martin.

Briefs

PHARMA LAYOFFS

Merck announced layoffs of about 12% of its workforce—6,800 employees and 400 positions to be left vacant—to be complete by 2011 in response to serious revenue shortfall from falling sales, approval delays, and patent expirations. The global restructuring should result in savings of $3.8 billion to $4.2 billion from 2008 to 2013, and comes on the heels of an earlier "restructuring" that laid off over 10,000. 40% of the jobs lost will be in the U.S., and 25% will be among mid- and senior-level executives. Research site in Japan, Italy, and Seattle will be closed by the end of 2009, but the company has offered some members of its molecular profiling group in Seattle relocation positions in Boston.

MERGER APPROVED

Invitrogen and **Applied Biosystems**' shareholders have voted to approve the two companies' $6.7 billion merger. More than 98% of the votes cast were in favor of the merger, and the two companies reported 80% of shares entitled to vote cast ballots. The merger has been approved by U.S. antitrust regulators, but is waiting for approval in the European Union. The merger was first announced in June.

CEO SWITCH

Affymetrix founder, CEO, and Chairman Stephen Fodor will become executive chairman of the company's board on January 1, and president Kevin King will become CEO and director. The transfer of responsibilities will take place over the next two months. King was hired in 2007 to take over the role of CEO. In his new role, Fodor will "identify and advance new platform technologies and explore new market opportunities for the company," said a company statement. Affy's third-quarter sales figures were down 21% year over year.

Riding Out the Storm

MICHAEL A. GREELEY

The economic environment is terrible—full stop. The financial devastation we are suffering with the echo-boom of the credit crisis will be felt for many years. We are a nation at war, the stock market is volatile, and we have the political indecision that inevitably settles in before fall elections.

Despite this backdrop, many venture capitalists felt relatively isolated from the credit market woes until recently. The credit crisis is now causing a much broader deterioration in the economic fundamentals. The situation for VC's is worsened by the overall lack of liquidity; there have effectively been no IPO's in a year and M&A activity is characterized by fewer, smaller deals—the negotiating leverage has clearly shifted to the acquirer. Even more troubling, the follow-on financing market, which many VC-backed companies rely upon to support product development and commercialization activities, is equally challenged.

It is also true though that many successful and valuable companies were launched in times like these. Notwithstanding the broad paralysis that has set in, there are strategies for entrepreneurs to deploy that will attract start-up capital. But they must be mindful of the context in which they operate.

Many successful and valuable companies were launched in times like these. There are strategies for entrepreneurs to deploy that will attract start-up capital.

Capital Efficiency

In October, the third quarter 2008 Silicon Valley Venture Capitalist Confidence Index was released, showing that VC sentiment has not been this low for many years—and it will likely continue to plunge. This has profound effects on the VC environment, as many firms will become inwardly focused on existing portfolio companies. In 2007 the broader U.S. venture industry raised over $30 billion; through the first half of 2008, the VC industry raised only $11 billion, and it will struggle to reach $20 billion by the end of this year. We are poised to see a dramatic repositioning of the VC industry.

The phrase of the times is "capital efficiency." Expect to see smaller financings—that is, the amount of capital raised in any given round should decrease (it is approximately $8 million per round now) as investors look to fund to nearer-term milestones. This will put a greater emphasis on management teams to be obsessively focused on building only those products customers will pay for now. Do not over-engineer product offerings with features that are not essential, hire only those people you need to accomplish near-term objectives. There will be a premium on successful managers who can demonstrate decisiveness and resourcefulness. Boards are counseling management teams to husband cash aggressively. In fact, many companies are exploring what the impact on business models would be if they were not able to raise any additional capital and how quickly the companies could get to breakeven.

Many venture firms are now discovering that they are under-reserved—the capital they had set aside for subsequent financing rounds is inadequate. Typically for every $1 invested, upwards of $3 to $4 should be reserved for later rounds. Many firms operated with lower reserves in the belief that other investors could be easily found to drive the additional rounds at higher valuations.

Expect also to see the "rolling thunder" of company closures wash across the landscape over the next 18-24 months, as companies fail to raise additional capital. This has a number of obvious negative impacts—in addition to investor distraction with closing portfolio companies, this will make many investors leery of investing in unproven business models and first-time entrepreneurs.

In 2007, according to Venture Economics, $30.7 billion was invested in venture-backed companies; through the first six months of 2008, $14.3 billion had been invested. Industry observers expect the third and fourth quarters fall off dramatically. The capital scarcity could last for the next couple of years.

All this demands that management teams be very thoughtful about the set of milestones that must be achieved in the short to medium to long terms. These would encompass product, market, and team milestones. Good investors will take the time and care to develop a financing strategy that overlays that.

Great CEOs recognize that it is not how much one raises but how much one owns of the company. Building a strong investor syndicate that is aligned on the long-term objectives but mindful of the steps in between, while perhaps more time intensive up front, should see a company through this challenging period.

Michael Greeley is a general partner at Flybridge Capital Partners and Chairman of the New England Venture Capital Association. He can be reached at michael@flybridge.com.

The Business of Stem Cells

RICKI LEWIS

Stem cell science is on the precipice of becoming big business. The potential markets for stem cell-based therapeutics are huge; in contrast to markets such as anti-neoplastics or anti-depressants, stem cell-therapy targets will extend into many fields of medicine. Already therapies are in various stages of development for Alzheimer disease, cancer, cardiovascular disease, diabetes, Parkinson disease, and spinal cord injury.

The hundreds of millions of potential patients helped to fuel the 2006 projection of an $8.5 billion global market by 2016 (calculated by Robin R. Young [Robin Young Consulting Group]). Market projections based on health statistics, however, cannot always factor in the effect of new enabling technologies or major discoveries, which may have been the case for human induced pluripotent stem (iPS) cells.

As 2007 dawned, the first iPS reports had just been announced and academic researchers immediately began to confirm and extend the groundbreaking work of Shinya Yamanaka (Kyoto University), for iPS cells are much easier to generate than embryonic stem (ES) cells. Pharma, just awakening to the potential of stem cells, at least in an active way, couldn't have helped but notice the new and exciting stem cell source. Given the facts that iPS cells may ultimately offer a much more expedient route to developing patient-specific cell therapies than ES cells, and that both restrictions and attitudes on all stem cell research are loosening, the $8.5 billion 2016 market projection is more likely an underestimate than an overestimate.

The hard questions are just now being asked. Will iPS cells stand in for human ES cells, once the bugs have been worked out, or will they find their own niche? What will companies do that have invested heavily in ES cells? For example, Geron (Menlo Park, Calif.) is pursuing the first human ES cell-based therapy, GRNOPC1, for spinal cord injury. They declined to comment on how the emergence of iPS cells is affecting the project.

Setbacks in stem cell research and commercialization have been rare, although one such setback was the May 2008 "clinical hold" that the FDA placed on Geron's trial of GRNOPC1. The company's stock dropped 20% in the wake of the announcement. Geron's news releases had indicated that the company expected to be the first to have an ES cell-based therapeutic approved—which could still happen.

Further Reading: *Stem Cells Come of Age*, by Ricki Lewis, Ph.D. Published by Insight Pharma Reports, November 2008. www.InsightPharmaReports.com

Pharma Comes on Board

Pharma emerged from the wait-and-see sidelines with the October 2007 debut of "Stem Cells for Safer Medicines" (SC4SM), an independent, not-for-profit partnership run by GlaxoSmithKline, AstraZeneca, and Roche, as well as a roster of U.K. government organizations. SC4SM will disperse funding to pursue use of human stem cells to develop lines of normal, human differentiated cells for high-throughput toxicology screening in drug development.

Pharma is supporting early stage academic research as well as partnering with nearer-to-market biotechs:

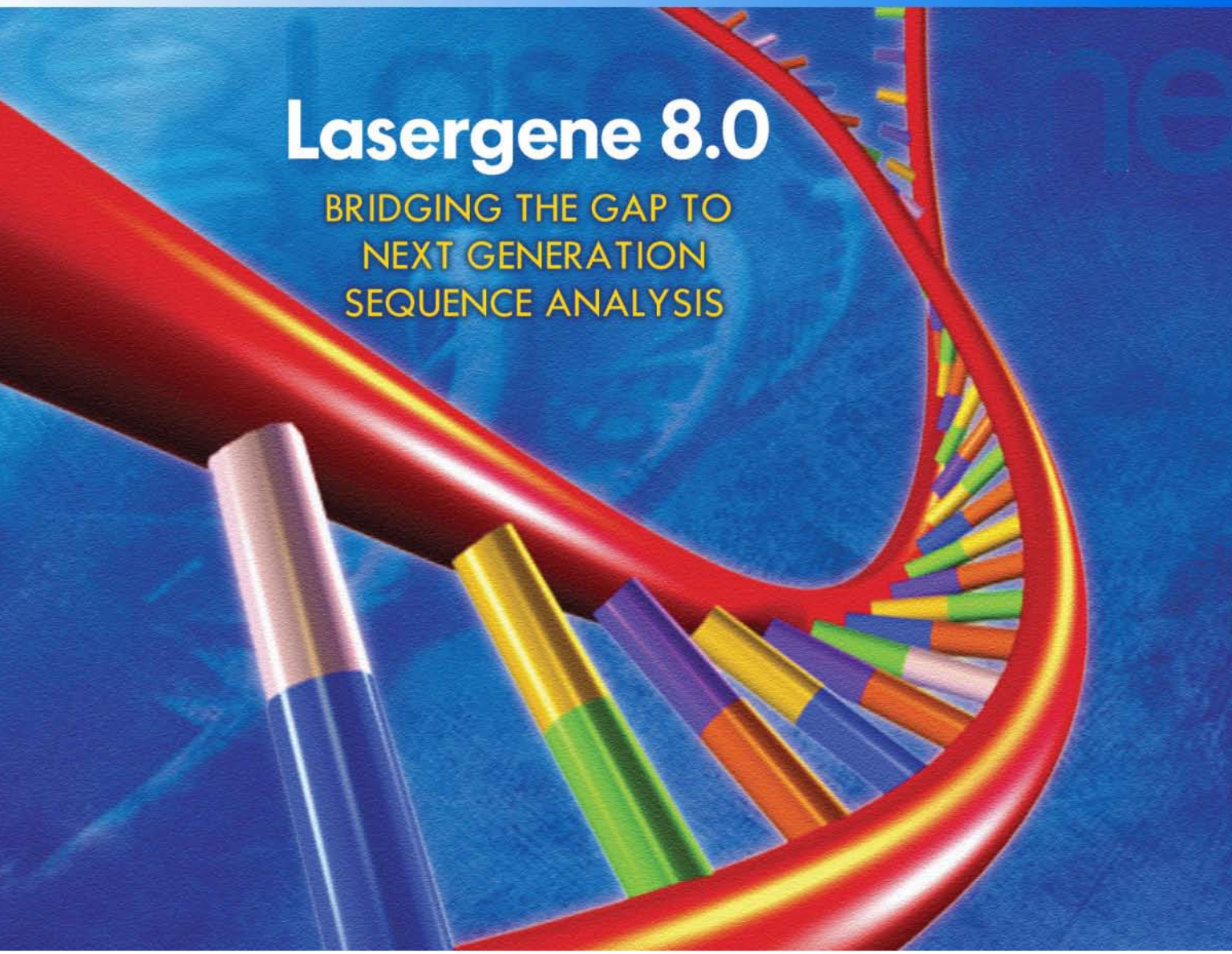
- GlaxoSmithKline has a five-year, $25 million+ collaborative agreement with the Harvard Stem Cell Institute to use all types of stem cells to develop new methods for drug screening. The funding will support research at Harvard University and several affiliated hospitals. GlaxoSmithKline supports biotech companies as well: It recently invested $1.4 billion in a project using antibodies to target cancer stem cells at OncoMed Pharmaceuticals (Redwood City, Calif.).
- Pfizer is founding San Diego-based EyeCyte to explore autologous stem cell treatments for disorders of the ocular vasculature. The effort builds on research from Martin Friedlander from the Scripps Research Institute. Friedlander approached Pfizer when he wanted to take his preclinical work toward the clinic. The investment begins with $3 million, with Pfizer retaining the option to buy the company.
- Johnson & Johnson Development Corp. is supporting Novocell (San Diego) in commercializing its platform to encapsulate human ES-derived pancreatic beta cells as a diabetes treatment. The proprietary biocompatible coating for the cells overcomes the problem of immune rejection of natural islets. Clinical trials are for long-standing type 1 diabetes patients.

The potential markets for stem cell-based therapeutics are huge; in contrast to some other markets, stem cell-therapy targets will extend into many fields.

The Harvard Stem Cell Institute, EyeCyte, and Novocell's novel islets provide a glimpse of what stem cells might do in the near future. Stem cells have developmental potential in that they can become many things; so, too, can their commercialization potentially help to treat an astounding variety of medical conditions.

Ricki Lewis can be reached at ralewis@nycap.rr.com.

DNASTAR®



Meet the needs of today with a solution for tomorrow.

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

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

New Lasergene 8.0 benefits your research with:

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

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

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

DNASTAR® Lasergene

Computational Biology

Grappling with Next-Gen Data Glut

Storage and data management drain resources even as sequencing speeds up.

BY KEVIN DAVIES

PROVIDENCE, RI—Vendors continue to release significant upgrades in their next-generation sequencing platforms, and the challenges of handling and analyzing all that data continues to frustrate users. CHI's Exploring Next-Generation Sequencing meeting* was complemented with a new track on data analysis to explore this issue.

The wealth of resources under development for next-gen sequencing was discussed by **Gabor Marth** (Boston College)—programs for base calling, read mapping, and SNP discovery to structural variation and data visualization. Marth said new analysis methods are needed for tailoring existing tools for different applications, improving analysis pipelines to focus on key results, and improving support for downstream analysis. The most popular software from Marth's own group includes the Mosaik Read Aligner and the EagleView genome assembly viewer.

"It's not the cost of storage, it's the cost of storage maintenance," said **Jacob Farmer** (CTO, Cambridge Computer). Farmer discussed various storage architectures, pointing out that it could take an organization two months to fully restore 90 terabytes (TB) of data from backup. He felt there is still room for tape when used in conventional backups and that a new concept of deduplicating data—for example in sequence data—is gaining traction.

Reece Hart (chair research computing, Genentech) said the demands of storing 300 TB of data and climbing cause a fair share of operational headaches, not least because of 1-2 hard drive failures a week. He disagreed with Farmer about the value of tape: besides being frail, Hart said it took six hours to restore 1 TB from backup. "Tape is really not a winning backup strategy for anything that is business critical." He favored mirroring storage, though hoped to move his current backup more than 20 miles away. Currently Genentech's biggest data source is high-content screening, but that will change as the company introduces its first next-gen sequencer in January 2009.

> "Tape is really not a winning backup strategy for anything that is business critical."
>
> **Reece Hart**, Genentech

During a lively panel discussion, representatives from several large genome centers took some heat from the audience for not releasing protocols and software. The Broad Institute's **Patrick Cahill** defended his center's efforts, saying it was working hard to release information, but it wasn't easy in such a fast-moving field.

More, More, More

Speakers from the big three commercial next-gen sequencing platforms all presented impressive improvements in throughput. **Jason Affourtit** (Roche/454) discussed the company's recently released Titanium series of reagents, which could deliver a fivefold genome coverage in two weeks compared to two months for James Watson's genome last year. The upgrades—including coating of wells to reduce crosstalk—result in longer read lengths of about 400 bases and run throughputs of about 500 million bases.

Not to be outdone, Illumina's **Gary Schroth** noted that a recent breakthrough in the deblocking chemistry of Solexa sequencing promised to extend individual DNA read lengths from 50 to 75 and possibly 100 bases in the next few months. As will be published in November, David Bentley and colleagues have sequenced the genome of an anonymous African male. The work generated 135 gigabases (Gb) data from 4 billion paired sequencing reads, identifying millions of SNPs and thousands of structural variants.

Applied Biosystems' **Michael Rhodes** said that his company has also sequenced the same African individual using its SOLiD system, generating 12-fold genome coverage so far. Analysis of the resulting data using a program called PolyPhen reveals hundreds of potentially deleterious mutations in coding genes.

Laurie Goodman, communications officer for the Beijing Genomics Institute (BGI) at Shenzen, provided a sneak peek of the first Asian genome sequence that will also be published this month. BGI has established itself as the third biggest sequencing center in the world, generating 21 Gb of sequence per day, primarily based on Illumina GA Analyzers. The Chinese researchers generated more than 3 billion 35-base reads, or 34-fold coverage of the Asian genome. The work, which was conducted last year, took two months using five instruments, and cost an estimated $500,000.

Patrice Milos, CSO, Helicos BioSciences, said Helicos is striving to increase throughput tenfold from 50 to 500 megabases per hour. The company has been improving its chemistry and extending average read lengths by testing on various bacterial genomes. Milos says Helicos intends to be sequencing human genomes in 2009. Helicos recently shipped its second HeliScope to Stanford University, where company co-founder Steve Quake is on the faculty. •

*CHI's Exploring Next-Generation Sequencing; Providence RI; September 23-25, 2008.

Editor's Note: CHI's Next-Generation Sequencing Applications conference will be held in San Diego, March 17-18, 2009.

Breathing New Life Into Old Drugs

ChanTest scours the Orange Book for drug hopefuls.

BY ALISSA POH

In drug development, it's not always out with the old and in with the new. Some companies feel that drug repurposing, while mostly still easier said than done, remains a viable alternative to the oft-financially crippling process of getting a new drug from pipeline to market.

Certainly, many biotech companies have made use of the FDA's electronic Orange Book, querying its "Disc" (Discontinued Drug Products) list, which contains thousands of drugs that made it through Phase I testing, only to be withdrawn for reasons other than safety: lack of efficacy, replacement with better drugs, commercial constraints, etc. Companies are constantly looking to "purposefully repurpose" this warehouse of abandoned drugs, finding new targets among protein families like kinases and G-protein-coupled receptors. But so far, only ChanTest Corp., based in Cleveland, is solely focused on reincarnating drugs by matching them with ion channel targets.

Now a decade old, ChanTest grew out of consulting work done by its founder and CEO, Arthur "Buzz" Brown, in the late 1980s for pharmaceutical companies on drug risk assessment related to cardiac ion channels. Brown's focus then was to find the target for Seldane (terfenadine), the first non-sedating antihistamine and also the "poster child" for non-cardiac drugs that caused sudden cardiac death.

"Seldane was a big breakthrough, because previous antihistamines crossed the blood brain barrier; people taking them often fell asleep at the wheel and died in car accidents," Brown remarks. But unexpected deaths from cardiac problems were noted among those on this new drug. "We showed that Seldane was directly blocking a particular potassium channel, hERG," Brown says, "and our cardiac risk assessment assay, which we then introduced, became very popular. One thing led to another, and ChanTest was formed."

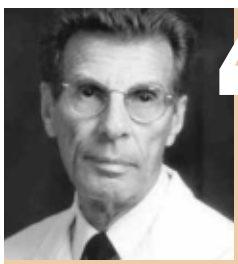
Channel Panels

Nowadays, ion channel testing for drug safety is a standard component of regulatory submissions at the FDA, prior to any drug being approved in humans. And ChanTest has an ongoing $10 million program aimed at developing the world's largest ion channel library, overexpressing important ion channels from the human genome in cell lines and validating each for its function. This library, Brown says, can be divvied up into "channel panels" according to tissue (e.g. cardiac, respiratory, CNS), therapeutic area (e.g. pain, seizure, diabetes/obesity), or ion channel family. The panels allow for primary and secondary pharmacodynamic profiling, where the first determines whether a particular channel is a drug target, and the second examines other channels said drug might hit, which could produce unwanted, off-target effects.

ChanTest has made use of the Orange Book to come up with a collection of approximately 3,000 drugs left by the wayside, and the firm's scientists are running these against their extensive ion channel library, searching for channel targets.

"We got the idea after investigating a drug for a client that, based on our hERG assay, should've been killing people," Brown explains. "However, the drug had been in man without such dire results; it also didn't have any efficacy for its indication, so it ended up being discontinued." ChanTest scientists then produced a hypothesis explaining why this drug wasn't lethal as expected, supporting the notion with their cardiac channel panel. And because the drug's profile (in a panel screen) looked intriguing, they decided to see if it would successfully terminate atrial fibrillation (AF) in an animal model. It did. Compound CT-1, as the drug is called, is currently undergoing a Phase 2a proof-of-concept clinical trial and looks "very promising," Brown says, with an AF termination rate of 60 percent, as opposed to 30 percent for present anti-arrhythmics.

ChanTest also offers fee-for-service drug repurposing to interested clients. Apart from the Orange Book, there are likely hundreds to thousands more abandoned drugs in the compound libraries of drug companies. "So a client says, 'I've got a hundred drugs I'm not using, but they're safe in man—think some might have an ion channel target?' and we go from there," Brown says.



> "Repurposing is not only a faster and cheaper way of getting drugs to the market; but thanks to genomics, we now have a lot more targets than we did before." **Arthur Brown**, ChanTest

ChanTest's foray into drug repurposing is still in the early days, Brown adds, so major success stories "are yet to come." But just because a drug fails against one target doesn't mean it can't be tried against another, he emphasizes, citing Viagra as a good example.

"Repurposing is not only a faster and cheaper way of getting drugs to the market; but thanks to genomics, we now have a lot more targets than we did before," Brown says. •

Personal  Genomics

KEEPING SCORE OF YOUR

Twelve months after the debut of the first personal genomics services, Kevin Davies expectorates and explores the pros, cons and controversies of "23 *et al.*"

ONE YEAR AGO, Iceland's deCODE Genetics kicked off a new era in personalized genetics with the launch of a direct-to-consumer service dubbed deCODEme. A few days later, California's 23andMe debuted its own service with the slogan: "Genetics just got personal. Don't worry. We're here to help."

For less than $1000, deCODE and 23andMe offered customers the exciting opportunity to gander their unique genetic quirks and glitches, as revealed by the identities of more than 500,000 DNA markers (single nucleotide polymorphisms, or SNPs). Such services would have been unthinkable just two years ago, before the explosion of genome-wide association studies (GWAS) began pinpointing susceptibility genes for common, complex diseases such as cancer, heart disease, diabetes, and mental illness.

Some of the attraction of these web-based services is undoubtedly recreational—what's my haplogroup? Do I have fast-twitch muscles? Am I a fast or slow caffeine metabolizer? Some people may be curious about tracing their family's roots or comparing genomes to friends and family members. Adoptees may be interested in gleaning something of their biological background. But these services, which now include offerings by Navigenics and SeqWright, translate the latest peer-reviewed genomic research into estimates of individualized disease risk, potentially allowing consumers the opportunity to confirm indications from their family history or revealing wholly unsuspected predispositions to a given disorder.

Sequence Secrets

The proselytizers of personal genomics acknowledge that these services are works in progress, but some anecdotal results have been remarkable.

Last summer, deCODE Genetics co-founder Jeffrey Gulcher, 48, reviewed his deCODEme data, which indicated he had a doubled lifetime risk for prostate cancer. Even though his PSA levels were fairly normal, Gulcher ended up getting a biopsy that revealed a grade 6 (Gleason scale) prostate carcinoma, which was successfully resected. "I'm not only the CSO, but I'm also a client," Gulcher joked at CHI's Beyond Genome conference last June. "This test may have saved his life," says deCODE CEO Kari Stefansson in all seriousness.

The co-founder of Navigenics, Los Angeles physician David Agus, received similarly alarming results when he took his firm's Health Compass analysis—an 80% risk of a heart attack. He told guests at the Navigenics launch party last Spring that he is taking appropriate lifestyle, dietary and medical action.

Sergey Brin, the co-founder of Google and husband of 23andMe co-founder Anne Wojcicki, recently disclosed that he

MARK GABRENYA

SEQUENCE

The author holding a mock Navigenics Health Compass scorecard that allows users to immediately assess their health risk profile. Conditions are grouped in columns according to predicted lifetime risk (<1%, 1–10%, etc.). Orange boxes denote conditions of higher than average risk to the individual.

Score Keepers: Stefansson, Avey & Wojcicki, Stephan, Agus

carries a mutation in a Parkinson's disease (PD) susceptibility gene called *LRRK2*. By searching all of his genotyped *LRRK2* SNPs through his 23andMe account, Brin learned that he carried the same point *G2019S* mutation as his mother, who has PD. This puts Brin, he says, in a unique and fortunate position: "I know early in my life something I am substantially predisposed to. I now have the opportunity to adjust my life to reduce those odds... and support research into this disease long before it may affect me... I have a better guess than almost anyone else for what ills may be mine—and I have decades to prepare for it."

A handful of anecdotal stories is no substitute for hard data of course, and the consumer genomics industry has sparked deep concern among the medical establishment and certain state regulatory agencies about the wisdom, validity, and legality of delivering medical information to consumers—and the public's limited ability to interpret their odds ratios and relative risks without expert assistance. Rather than fork out $1000 for a genome scan, the *New England Journal of Medicine* argued earlier this year that it made more sense to buy "a gym membership or a personal trainer" and implement dietary changes to reduce the risk of heart disease and diabetes.

Last summer, the State of California issued "cease and desist" letters to 13 consumer genomics providers, but while some firms retreated, the major providers satisfied the State that their tests are performed in CLIA-certified labs with appropriate physician oversight.

Regulatory rumblings notwithstanding, it is easy to overlook just what has been achieved in the past 12 months thanks to the collective efforts of what BioTeam's Michael Cariaso dubs "23 *et al.*" These genome scans are not, the providers stress, *diagnostic* tests, but a means of providing information that mirrors the stunning progress researchers are making in dissecting the genes underlying complex traits. Indeed for most genes, the individual impact on a person's relative disease risk is minor. One exception is Apolipoprotein E (ApoE), which was strongly associated with late-onset Alzheimer's disease (AD) by Duke University professor Allen Roses and colleagues (see, "A Virtual Pharma Organization," *Bio•IT World*, June 2008). This is the gene that James Watson had redacted from his genome sequence last year rather than learn his potential AD risk.

"The odds ratio of ApoE4 is 3.0, which is pretty substantive," explains Boston University's Robert Green, who is conducting a trial to study reactions to Alzheimer's genetic testing. By contrast, most gene associations in complex diseases produce odds ratios of 1.1 or 1.2 and thus have limited significance by themselves. "But people who are not statistically sophisticated will not necessarily understand that. All they'll see is that the risk is higher or lower," Green warns.

Green talks of a "nightmare scenario" in which a woman is sophisticated enough to order a personal genome scan, but unable to distinguish the *BRCA1* hereditary breast cancer gene from the breast cancer markers offered by 23 *et al.* "Oh my God, I got the gene for breast cancer," she thinks. Her surgeon studies the printout and concurs: "Yeah, you're positive for the gene, I guess we better [schedule surgery]."

Running the Bases

During 2008, I signed up for personal genomics services from four providers: deCODE, Navigenics, 23andMe and Houston-based SeqWright. The firms' philosophies may vary but the procedures are similar. The consumer's DNA is surveyed on a commercial or customized Affymetrix or Illumina DNA chip for 500,000 to 1 million SNPs. The results

PERSONAL GENOMICS COMPANIES

	23andMe	deCODEme	Navigenics	SeqWright
Launch	Nov-07	Nov-07	Apr-08	Jan-08
Platform	Illumina	Illumina	Affymetrix	Affymetrix
Conditions/Traits	91	31	23	16
List Cost	$399	$985	$2500 + $250 annual sub	$998
Counselor	No	Referrals	On staff	No
Pros	· Social networking · Price	· Generates primary data · Physician marketing	· Actionable conditions · Medical outreach	GLP compliant
Cons	· "frivolous" image	· Financial health	· Expense	· Minimal marketing

are delivered via the web and (with one exception) regularly updated as new findings are published. These services are now as cheap as $399, thanks to 23andMe's recent price cut.

Navigenics: Health Compass

At $2500, the Navigenics service is the most expensive whole-genome SNP service on the market—excluding Knome's six-figure sequencing service. Navigenics is working closely with the medical community and does not consider itself simply a direct-to-consumer offering. For example, it is collaborating with Scripps, Affymetrix, and Microsoft in a 20-year study to determine if personal genomic testing encourages participants to make beneficial lifestyle changes.

Last Spring, I visited Navigenics headquarters in Mountain View, Calif., to discuss my results with Elissa Levin, the director of the company's in-house, on-call team of genetic counselors. Navigenics is the only company that offers this direct resource. "I think people would benefit from having some expert consultation," of their genome scans, says Green, noting that the family physician isn't necessarily qualified for that role.

There's nothing recreational about the Navigenics Health Compass: it won't tell me whether I can smell asparagus in my urine or whether my maternal ancestors hail from Doggerland. The focus is strictly on medical conditions—in particular, *actionable* conditions for which Navigenics believes there is a beneficial medical or lifestyle intervention. "If we can detect Alzheimer's at age 21, we open up 50 years—I call it the therapeutic window," explains co-founder and CSO Dietrich Stephan. "As opposed to presenting when half your brain has been melted by disease, you present when you're totally cognitively normal."

The Health Compass currently presents information on 21 diseases. The results are presented clearly and intuitively: conditions are arrayed in columns according to estimated lifetime risk, from less than 1 percent to above 50 percent. This reminds clients that their lifetime risk of a heart attack, although perhaps less than the general population, is still greater than say restless leg syndrome, even if their individual risk for that disease is increased. Disease boxes shaded orange indicate a significantly elevated individual risk. I quickly survey only five orange boxes—a good omen perhaps.

The single most predictive result is for AD. For the 3 percent of the population that carries two copies of the ApoE4 allele, the risk of AD is increased ~ 15-fold, and my curiosity is heightened because the disease stole my great aunt Gwyneth. The silver panel puts my lifetime risk at 4.5 percent—half the national average.

There is similarly reassuring news for lupus, age-related macular degeneration, multiple sclerosis and type 2 diabetes. For each disease, I can learn my precise genotype and a wealth of medical information. I can also print out a thick report to take to my physician. But some conditions do bear further inspection: my risk of heart disease is no better than the national average, about 40 percent, and my risk of prostate cancer is slightly elevated. In most cases, these risks are calculated on just a few known genes, with the likelihood that many more genes remain to be detected, whether by GWAS or resequencing for rarer variants. My calculated risk will be revised as each new variant impacts the overall likelihood of that disorder.

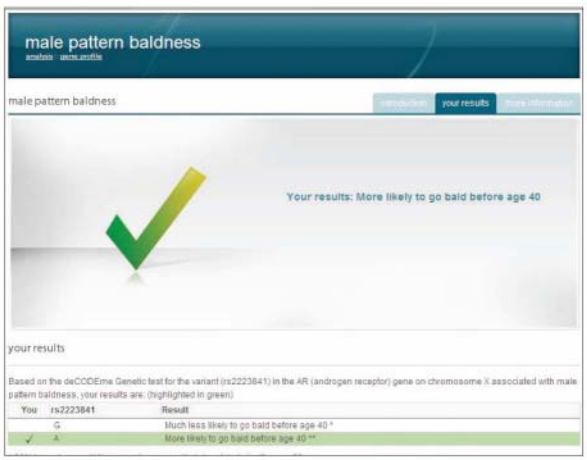
One final entry catches my eye: obesity is in an orange box, my lifetime risk pegged at the national average of 34 percent, because I carry a risk allele for the "fatso" gene, *FTO*. At 6'2" and barely 165 pounds, with a limited exercise regimen consisting of refereeing a few high-school soccer matches, that isn't something I will worry about.

deCODEme

While deCODE Genetics is best known as a drug development and diagnostics company, Kari Stefansson says "The original business proposal was always to use genetics for preventative health care. The idea of marketing direct to consumer is something we've been toying with for several years."

Unlike its competitors, deCODE is driving much of the original research in complex traits (see, "Assessing SNP Associations"). Stefansson says Navigenics and 23andMe are basically "dot com companies... They are mostly marketing discoveries that we have made." In October, deCODE scientists published two SNPs associated with skin cancer (basal cell carcinoma), and introduced those SNPs into the deCODEme service. But any advantage was short-lived: within 24 hours, 23andMe reminded its users that they could search for the same SNPs in their raw data. Although competitors can use deCODE's diagnostic markers "with impunity," Stefansson warns "that is not going to last long. Pretty soon, these patents are going to issue."

The deCODEme service offers information on more than 30 conditions. It



DeCODEme delivers the bad news: you're going bald.

The deCODEme genome browser lets users survey their SNPs on the latest reference genome.

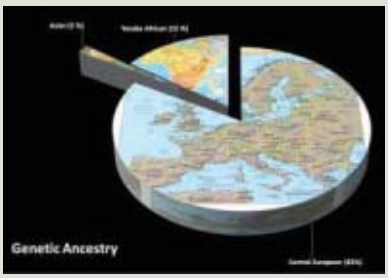
includes information on ancestry and some personal traits, including hair and eye color and—unfortunately in my case—baldness. More surprisingly, I'm an asymptomatic carrier for hemochromatosis (ironic, as I'd published the paper on the cloning of that gene a decade ago). There's also a tool to compare my genome to dozens of people of different ethnic ancestry. This tells me I am genetically more similar to Craig Venter than, say, an Mbuti pygmy—a feature some consumers might deem "recreational," others not so much. A more useful tool is deCODE's gorgeous java-fueled genome browser: using this, I can pinpoint all of my SNPs matched against the reference genome, zooming into the raw sequence if desired.

23andMe: Gene Journal

Not surprisingly, 23andMe, co-founded by Linda Avey and Anne Wojcicki, has enjoyed lavish media attention for the past 12 months. That's attributable to the Google connection, the celebrity spit parties, and non-stop appearances from the *Today Show* and New York Fashion Week to the World Economic Forum and even the Bio-IT World Expo (see, "23andWe," *Bio•IT World*, June 2008). The company provides the deepest dive into your genome data, presenting information on ancestry, traits (earwax, heroin addiction), carrier status for Mendelian disorders, and risks for common diseases.

The publicity surrounding character traits earned 23andMe a "frivolous" tag in some quarters that Avey and Wojcicki reject. "We really applaud the fact that people are interested in their genomes for reasons other than just their health," says Avey. "The ancestry work we do—I wouldn't call that frivolous at all."

In September 2008, 23andMe slashed the price of its service from $999 to a mere $399, using a second-generation Illumina platform. The Gene Journal assigns one to four stars for more than 90 conditions, based on the firm's internal assessment of the validity of the original research. I learn, for example, that I'm not a carrier for the most common cystic fibrosis mutation. But there are differences with reports from other services—for example, 23andMe calculates I have an increased risk of macular degeneration.



The author's ancestry painting by 23andMe (above) is a boring solid blue (100% European). By contrast, SeqWright (below) points to a small contribution of Asian and African DNA.

The medical reach of the Gene Journal is illustrated not only by Brin but also by his wife. Wojcicki has revealed that she is a carrier for Bloom's syndrome, a recessive disease that, like Tay Sachs and a dozen or so other conditions, is more frequent among people of Ashkenazi Jewish descent. For Wojcicki and her husband, such information is anything but recreational: they are expecting their first child.

The ability to search the full SNP dataset is also valuable. For instance, a recent *New England Journal* report identified a SNP that predisposes towards statin-associated myopathy. It was trivial to key in the identity of that SNP and learn I did not carry that allele.

The ancestry painting involves a proprietary algorithm, explains senior director of research Joanna Mountain, to compare groups of SNPs against those catalogued in the HapMap project. Other services had said my genome contained traces of African or Asian genetic material, but the 23andMe report is unequivocal and consistent with my own self identity: my chromosomes are 100% blue, or European. Avey, who has the same pattern, called her chromosome painting "very boring."

23andMe has pioneered the social networking aspect of its service, encouraging customers to compare genes to friends and family and join a community effort called "Research 2.0."

SeqWright: GPS

Based in Houston, SeqWright's Genome Profiling Service (GPS) has received scant public attention compared to its competitors since its launch in early 2008. The company, co-founded by Baylor genome center director Richard Gibbs, is a genome service contract research organization, but it already possessed a CLIA-certified microarray facility.

SeqWright's GPS lacks some of the bells and whistles of its competitors. My report was delivered in May, but it is not dynamically updated. The report is presented as a single HTML page—mine presented 16 medical conditions. It conveys odds ratios for each tested SNP, but the GPS doesn't offer an aggregate likelihood ratio or lifetime risk for each disease. When presented with a list of 17 SNPs for Crohn's disease or 26 for type 2 diabetes, it's unclear for a geneticist what to make of the information, let alone a lay person. Nor does it list the genes that contain each SNP, citations to the original literature, or the ability to dynamically search raw SNP data.

Digging Deep

If you are not content to wait for the personal genomics companies to decide which conditions are officially presented, then Michael Cariaso's Promethease program may be just the ticket. "The pure computer programmer in me has always wanted to write Promethease," says Cariaso. "I want to have my own DNA as data and figure out what the hell it means."

After downloading my raw SNP data,

Assessing SNP Associations

Before including a disease SNP in their consumer genomics offerings, 23 *et al*. employ teams of very smart geneticists and bioinformaticians to assess the merits of every peer-reviewed association. Many of these papers—and the SNPs genotyped as a result—stem from deCODE's ground-breaking research on complex diseases. "We have isolated 50-70% of the new discoveries," in the past two years, says Stefansson, who ticks off precedents in type 2 diabetes, heart attack, and prostate cancer, to name a few. "We aren't marketing tests unless they have a greater effect on the risk of disease than elevated cholesterol."

At Navigenics, Michele Cargill, who has the distinction of having worked with both Eric Lander and Craig Venter, leads a group of more than half-a-dozen geneticists who weigh every reported gene association in the top-tier journals. Not only must the association pass muster statistically, the disease must be actionable. If a robust gene association for Lou Gehrig's disease (ALS) should materialize next month, Navigenics won't include it unless it believes there is a genuine medical or lifestyle change that can benefit the individual.

Stephan says his colleagues perform a "decision heuristic," a laundry list of criteria that includes:

- The genome-wide association study must include at least 500,000 evenly spaced SNPs.
- The study must be replicated in at least two independent populations.
- The report must be controlled for ethnicity, gender, age, etc.
- Ideally there is some functional evidence surrounding the association.

At 23andMe, a similarly qualified team of geneticists, many of them Ph.D.s formerly with Stanford, review papers as they appear in press. The Gene Journal features many more traits and conditions, but like newspaper movie reviews, each condition is rated one to four stars. Only a four-star rating is deemed "established research." In one case, a result published in *Science* earned a single star.

Most of my own results across the three major platforms that offer aggregate likelihood risks were in close agreement, but there were some intriguing disagreements. These could arise for a number of reasons, including different platforms (and SNPs), statistical programs and criteria. For example:

- 23andMe reports a 50% increased risk for age-related macular degeneration based on three SNPs, but Navigenics and deCODEme indicate a lower than average risk.
- Navigenics reports a reduced risk of rheumatoid arthritis, contrary to deCODE and 23andMe.
- There is a wide discrepancy between the reported lifetime risks of abdominal aneurysm between deCODEme (17%) and Navigenics (3.1%).

I ran Promethease, which annotates my genotypes based on information in SNPedia, a wiki for SNP data built by Cariaso and Greg Lennon. This reveals hints of associations with many other conditions, although this is very preliminary data. Cariaso recently demonstrated the potential of his program when he won a 23andMe-sponsored competition to predict the traits and characteristics of 23andMe's anonymous demo account holder, based solely on a handful of SNPs.

All of the consumer genomics firms have expressed interest in shifting to whole-genome sequencing once the price is right. Stephan says Navigenics will introduce some form of exome sequencing in 2009. Knome CEO Jorge Conde says he expects to have sequenced 20 individuals by the end of 2008. The sequencing is performed by the Beijing Genomics Institute in Shenzen; clients fly to Boston for a daylong debriefing before receiving their data on a USB thumb drive.

Knowing one's full genome sequence won't be a godsend. There are still glaring deficiencies in our ability to interpret genome data or predict the significance of the thousands of novel SNPs and copy number variants unearthed in each new human genome. After originally being told last year he carried a mutation in the *BRCA1* breast cancer gene, James Watson later learned it was nothing more than a "benign Irish polymorphism." That may not concern an 80-year-old man, but such a misinterpretation error could have catastrophic results for a young woman with a family history of the disease.

Yet despite legitimate concerns about the validity and significance of every reported gene association, Stefansson has no doubt where this is going: "Within the next 3-5 years, pretty much every college-educated person in the U.S. is going to have a profile similar to the one found in deCODEme or 23andMe or Navigenics."

Navigenics' Stephan goes further. "Ultimately," he says, "[the goal] will be: sequence the genome, put it in a big computer, push a button, and get a rank order list of things you are at risk for. It will supplant newborn screening and all molecular diagnostics."

This will surely hasten some sort of federal regulation, which would be widely welcomed. The consumer genomics trend is "absolutely unstoppable," says Green, who adds: "The question isn't really, should it be happening or can we stop it? The question is to foster an incremental path along which it's done better and better."

And let's not forget the environment. Two years ago, my soccer days came to an abrupt end when my face latched onto a thunderous clearance at close range, resulting in a detached retina that no genome scan could have seen coming. •

Editor's Note: The author thanks deCODE Genetics, Navigenics and SeqWright for providing complementary personal genome services.

Triple Play

Tripos' collaboration with Wyeth enhances research and workflow efficiency.

BY ALISSA POH

If there's one word Big Pharma loves, it's "streamlining." So Tripos International's latest software solution—developed as a joint venture with Wyeth and consulting firm Accenture—could be on other companies' shopping lists pretty soon.

"It's a system by which companies with a lot of research data can create an integrated, real-time platform for accessing, analyzing, and sharing all that information," says Patrick Flanagan, Tripos' general manager. "Prior to this, folks at Wyeth would spend a significant portion of their time collecting and aggregating data on very discrete topics. Now they claim they're doing analyses they never would've done previously, as a result of time-saving and access to information they didn't have before."

As far as Tripos is concerned, this collaboration is the "evolution," as Flanagan puts it, of its Benchware Discovery 360 product, first developed several years ago when Bristol-Myers Squibb (BMS) needed to reduce operational costs by streamlining its drug discovery processes. Discovery 360 has been described as "the first commercially available integrated discovery environment" that resides atop a company's data repositories, giving scientists and project managers alike a single point of entry by which to consolidate structural, biological, and chemical data.

Wyeth picked Tripos out of several competing software providers, Flanagan says, to leverage and extend the capabilities of Discovery 360 for its *Next-Generation Discovery IT* initiative.

"It was a fairly deliberate process," says Steve Howes, senior director of bioinformatics at Wyeth, of the selection. "We wanted people familiar with the [software] marketplace, but more importantly, who were founded on science and actually had experience working with scientists. So after surveying the vendor landscape and reviewing several options, we chose Tripos because they met both criteria."

Potent Combination

"There are approximately 1,000 discovery scientists at Wyeth, including both biologists and chemists, who can get value from this capability," Flanagan adds. "Wyeth wanted them to exist comfortably in this large environment, and to have a common set of tools for conducting experiments *in silico*." Tripos already had the foundation of such workflow-improving tools at hand—namely Discovery 360—and was charged with collaborating to build an "enterprise-level" solution that would tie together Wyeth's different, disparately-located databases. Accenture developed a near real-time data warehouse accessed by Discovery 360.

The resulting *NextGen* initiative utilized the technology first used with BMS (called SmartIdea), with some significant additions to meet the needs of the Wyeth scientists in its final design, deployed earlier this year. All in all, a "pretty potent combination," Flanagan says.

> Biologists have access to chemists, team leaders have full visibility of their team's experiments.

"It comes down to this, really—biologists [at Wyeth] now have access to chemists and vice versa, team leaders have full visibility of their team's experiments, and management folks can view and impact the full R&D portfolio," he elaborates.

Not that finalizing this venture has been without its challenges, of course—in particular what Flanagan calls the "data harmonizing piece." Wyeth, like most large pharmaceutical organizations, had a variety of different databases, software, coding systems, and processes; producing a common information-processing format [within Wyeth] for Discovery 360 thus required "a large team effort," says Howes, adding that it was a great partnership with talented people on all sides.

Currently, 550 of Wyeth's discovery scientists have been trained to use the new system, in four U.S. sites, and Wyeth's goal is for it to be deployed to all 1,000 by year's end. "Now that it's in place, we're very pleased at the feedback we're getting from our scientists, and look forward to expanding it through discovery, making it easier to do science," Howes says.

Since announcing the solution in August, Tripos has been involved in "conversations of varying degrees of sophistication," Flanagan says, to offer this service to several other potential clients. "A couple of the places we're talking to are organizations with [homegrown] systems, but they've recognized the burden of owning and maintaining it in today's macroeconomic environment, and they're leaning toward having a vendor like us take matters in hand."

Tripos also hopes to offer smaller biotechs the ability to plug into Discovery 360-like systems in the near future. "We want to make it easier for organizations lacking the muscle and budget of a large pharmaceutical firm to also access this type of resource," Flanagan says.

As Tripos' acquisition of Pharsight Corporation, a company providing PK/PD modeling and simulation tools, is finalized, Flanagan sees the latter's expertise in optimizing preclinical and clinical drug development as an added bonus that could potentially be integrated into future versions of Discovery 360.

"Our view [at Tripos] is that chemical informatics and modeling software companies can have the most value over time along the entire R&D spectrum," he says. "Our focus is to acquire others complementary to us, versus duplicative, in terms of what they offer to the public. Pharsight is our first acquisition along this template, and it should be a great extension into an area where we don't [currently] have a presence." •

Clarix Joins Phase Forward

Purchase joins EDC and interactive response technology.

BY DEBORAH BORFITZ

Responding to intense interest among study sponsors and investigative sites for consolidated technologies, Phase Forward invested $40 million in cash to acquire privately held Clarix LLC in early September. The move tightly integrates Phase Forward's flagship InForm electronic data capture (EDC) system with Clarix's interactive response technology (IRT), which has done more than its competitors to augment the traditional telephone interface with the power of the Web.

So says Phase Forward vice president Martin Young, noting that the company has worked with more than 10 different IRT providers across roughly 100 clinical trials. That Clarix has a fully Web-integrated model for subject randomization, predictive medication inventory management, and operational management and reporting is only one of its chief attributes. Clarix also allows clinical teams to build their own ad hoc reports during the course of a trial and not just use the standard, preconfigured ones.

From contracting to service delivery, customers want to deal with a single project manager and help desk as they employ more technology in the clinical trials process, says Young. They're also interested in speeding up study start-up time. With Clarix IRT, trials can be set up in less than half the time it typically takes other IRT providers.

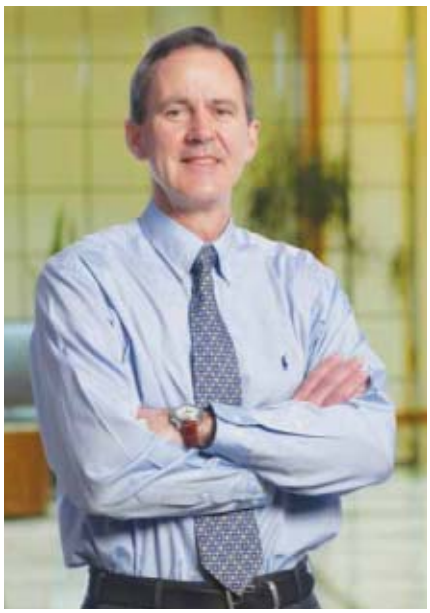
Being acquired was not "top of mind" for Clarix, says CEO and founder Jagath Wanninayake. "We were self-funded from the get-go and set out to build a sustainable, long-term business." Although Clarix has turned down other acquisition opportunities, the deal with Phase Forward made sense because the companies are "like minded" in terms of enabling more efficient clinical trials.

One of the biggest hurdles facing the newly combined companies is keeping up with the "insatiable desire" of the Phase Forward sales team for information about the Clarix IRT solution, says Young. For now, a bi-directional interface will link Clarix to InForm as requested by clients as they set up new studies. Clarix will "effectively be a part of" Phase Forward's eClinical suite.

Synergistic Sell

"It's a synergistic sell to InForm's customer base," which consists of 280 sponsor companies and another 100 clinical research organizations, says Young. Clarix will also continue to be sold separately. Currently, Clarix has more than 30 customers, including three of the top five and 11 of the top 25 pharmaceutical companies. Its IRT system has been translated into 55 different languages and dialects and has more than 45,000 users in 62 countries.

Clarix generated approximately $2.7 million in revenue in 2007 and is expected to contribute that same amount to Phase Forward's results for the remaining four months of 2008, says Young. Clarix is projected to produce about $6.3 million for the full year 2008 on a stand-alone basis. In 2009, the company could be contributing $13 to $15 million to Phase Forward's coffers.



Martin Young

"Growth has accelerated over the last several years as a reflection of adoption [of IRT] in the marketplace and in recognition of the differences we have to offer," says Wanninayake. The design focus of Clarix IRT is encapsulated by its motto: "It's our system, but it's your data."

Clarix will continue to operate from its existing offices in Radnor, PA. •

Double Layer of Semantics

BRIDG aims to synchronize CDISC and HL7 semantics.

BY DEBORAH BORFITZ

Subject matter experts and information technology professionals involved in clinical trials should both be happier with the next release of BRIDG (Biomedical Research Integrated Domain Group), the domain analysis model aimed at synchronizing semantics within the clinical research and health care enterprises. Semantically speaking, BRIDG will be taking a "two-layer approach" that incorporates accepted clinical terminology and definitions as well as technical representations around which software programs and messages can be written.

So says Julie Evans, senior director of technical services at the Clinical Data Interchange Standards Consortium (CDISC). Five years ago, CDISC initiated construction of BRIDG to support

Clinical Research

harmonization of its own set of standards as well as to "interface more cleanly" with Health Level Seven (HL7), the standards-settings organization for electronic interchange of clinical, financial, and administrative information among health care-oriented computer systems. Since then, the National Cancer Institute, Food and Drug Administration, and HL7's Regulated Clinical Research Information Management (RCRIM) Working Group have joined the effort.

BRIDG is currently a "mix" of technical and domain expert semantics, says Evans. But an alpha version of the two-layer model is expected to be available later this fall, improving the ability of all user types to understand and use BRIDG. There will be mapping between the two levels of semantics that will, for example, allow computers to unambiguously process words describing such concepts as vital signs and adverse events in a semantically interoperable way.

The only significant implementation of BRIDG at the moment is the NCI's Cancer Biomedical Informatics Grid (caBIG) initiative, an information network enabling all constituencies in the cancer community—researchers, physicians, and patients—to share data and knowledge. The data analysis model was adopted in 2005 by RCRIM, which is working toward creation of BRIDG-based messages for the interchange of data within and among trial-sponsoring companies and the FDA, says Evans.

Last year, the FDA included BRIDG in its five-year IT plan as a foundation for several projects. Requirements about how sponsor data is to be sent to the FDA is being negotiated among BRIDG stakeholders, says Evans. One of the goals of BRIDG is to achieve computable semantic interoperability in protocol-driven research and health care, so that computer systems precisely understand terms and their meaning and can thus effectively communicate, says Evans. The other goal is to allow clinical trial data to be connected to electronic health record (EHR) data. Substantial progress has been made on the first goal, in that nearly all standard clinical trial information such as patient demographics and other safety information have been semantically described. Disease-specific terms are not included and it has yet to be determined which ones will be. "Even with what we have so far, the scope is huge," says Evans.

BRIDG follows the HL7 Development Framework because it's an accepted, stepwise way to develop HL7 messages, says Evans. "There is no other framework for this, other than generic ones for software development" whose vernacular is not specific to HL7 message development.

CDISC is also involved in solutions to such hurdles as the acceptance of digital signatures by the CDISC Operational Data Model and electronic capture of trial data directly from EHRs. •

[GUEST COMMENTARY]

Flexibility in Translational Research Informatics

Tools are becoming increasingly crucial to integrate and analyze translational data.

BY JONATHAN SHELDON

The growth of multi-site, multi-technology translational research studies (the clinical application of scientific medical research from the lab to the bedside) is placing increased emphasis on informatics capabilities to integrate, analyze, and visualize translational data. Many commonalities are emerging in translational studies, both in medical institutes and drug companies, suggesting that the time is right to look at a more product-based approach to translational informatics. While configurability is still necessary for any system, it is now a considerably reduced part of the total solution compared to earlier custom-made systems.

Although restricted by HIPAA compliance and patient security issues, medical institutes have been blazing a trail for pharmaceutical and biotech companies in their adoption of translational research. Many large research institutes are making great progress in implementing electronic, searchable patient databases. This allows them to leverage their extensive patient information and samples to further understand the mechanisms of disease and characteristics of their patient populations.

By contrast, pharmaceutical and biotech companies have historically been slow in fully adopting translational research. This is often due to data access restrictions and reflects the view that clinical data is for support of FDA submission only, not for research purposes. With any delays costing the company upwards of $1 million per day, "nice to have" data was typically not an option. However, this is beginning to change. More and more companies are seeking information and proposals about tools to support translational research. Interest in an informatics infrastructure to support translational research is increasing, particularly for the integration of biomarker data, hypothesis generation, and ad hoc querying.

These fundamental changes are cited as a means to better utilize biomarker techniques and existing and public clinical data sets. It is also a move away from rigid use of clinical data and toward a model that supports an iterative approach to compound development. In this model, compound progression only occurs in combination with treatment selection and safety biomarkers, and a solid understanding of disease pathophysiology. This change enables drug companies to move away from linear processes toward a business model that produces safer, more effective drugs with lower development costs.

> Translational research is a move toward an iterative approach to compound development.

Mash-ups and Data Marts

Hypothesis generation and ad hoc querying for translational research requires the ability to flexibly integrate, analyze, and visualize data. In Web 2.0 terminology, this is a data "mash-up"—taking information from data warehouses and marts, web services, and personal data tables such as Excel, and providing a dynamic view of the data to address a particular scientific question. Mash-ups sit alongside (rather than replace) typical data warehousing approaches to enable fast responses to business questions and flexibility in hypothesis generation. This type of configurable system can also take advantage of new services and databases from programs such as caBIG and I2B2. The provision of web services and data schemas can be easily integrated into data mash-ups. Users can also feed data from legacy systems using this approach.

Another approach to data integration in translational research is using disease-specific data marts. By extracting de-identified patient data from more operational focused databases and sample status from LIMS sources, information about patients and sample status can be combined into a common view. This approach enables the use of disease ontologies and Extract Transform and Load (ETL) operations to take legacy data and make it available to new web-friendly architectures.

Decision Support for Clinicians

With the growing availability of comprehensive clinical data, researchers can select multiple cohort studies by slicing and dicing patient populations to identify the most appropriate biomarkers. New tools are now available to enable researchers and clinicians to browse clinical and integrated sample data to select disease and non-disease cohorts based on multiple criteria. Critically, these systems are designed in collaboration with clinicians to ensure appropriate interaction with and display of accurate predictive data that supports evidence based decision making.

The combination of these capabilities and trends is leading the way to a more configurable, product-based approach to translational research informatics, which takes data from multiple sources and delivers it to end users in a meaningful way. By working together, medical institutes and drug companies can learn from each other to re-engineer the process of drug discovery and deliver a major step forward in understanding the molecular basis of disease. •

Jonathan Sheldon is the CSO of Inforsense. He can be reached at jsheldon@inforsense.com.

Cambridge Healthtech Institute's Seventh Annual

Bio·IT World

CONFERENCE & EXPO '09

April 27 – 29, 2009 • World Trade Center • Boston, MA

Enabling Technology. Leveraging Data. Transforming Medicine.

KEYNOTE PANEL:

The Future of Personal Genomics:

A special plenary panel discussion featuring

Robert C. Green, M.D., M.P.H. Professor of Neurology, Genetics and Epidemiology, Boston University Schools of Medicine and Public Health

John Halamka, MD, MS, CIO, Harvard Medical School

Clifford Reid, Ph.D., Chairman, President and Chief Executive Officer, Complete Genomics

Dietrich Stephan, Ph.D., Co-founder and Chief Science Officer, Navigenics, Inc.

KEYNOTE PRESENTATIONS BY:

Chris Dagdigian, Founding Partner and Director of Technology, BioTeam, Inc.

Clifford Reid, Ph.D., Chairman, President and Chief Executive Officer, Complete Genomics

CONCURRENT TRACKS:

Track1: IT Infrastructure & Operations

Track 2: Informatics for Genomic Medicine

Track 3: Predictive and *in silico* Science

Track 4: Computational Chemistry

Track 5: Clinical and Medical Informatics

Track 6: eHealth Solutions ***NEW***

Platinum Sponsors:

BLUE ARC

ISILON SYSTEMS

Gold Sponsors:

accelrys

INGENUITY SYSTEMS

Bronze Sponsor:

BIOTEAM Enabling Science

Official Publication:

Bio·IT World

Please mention keycode U02 when registering!

Organized & Managed by:

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

www.Bio-ITWorldExpo.com

CONFIRMED SPEAKERS

Ingrid Akerblom, Ph.D., Executive Director, Clinical IT, Merck & Co., Inc.

Samuel (Sandy) Aronson, Executive Director of IT, Center for Personalized Genetic Medicine, Partners HealthCare, Inc.

Andreas Bender, Ph.D., Assistant Professor, Medicinal Chemistry Division, Leiden / Amsterdam Center for Drug Research

Brian Bissett, MBA & MSEE, Staff Scientist, Molecular Properties, Pfizer

Mark S. Boguski, M.D., Ph.D., Founder, Resounding Health Incorporated

Barry A. Bunin, PhD, CEO & President, Collaborative Drug Discovery, Inc.

Chris Dagdigian, Founding Partner and Director of Technology, BioTeam, Inc.

Derek Debe, Ph.D., Senior Group Leader, Scientific Informatics and Automation, Abbott Laboratories

David DeCaprio, Ph.D., Associate Platform Director for Chemical Biology and Novel Therapeutics, Broad Institute

Robert C. Green, M.D., M.P.H. Professor of Neurology, Genetics and Epidemiology, Boston University Schools of Medicine and Public Health

Rajarshi Guha, Ph.D., Visiting Assistant Professor, Informatics, Indiana University

John Halamka, M.D., M.S., CIO, Harvard Medical School

Anton Hopfinger, Ph.D., Professor, University of New Mexico, Editor, Journal of Chemical Information and Modeling

Andrew L. Hopkins, D.Phil., FRSC, SULSA Research Professor, Translational Biology & Chair, Medicinal Informatics Division, University of Dundee

Jeremy L. Jenkins, Ph.D., Research Investigator, Lead Discovery Informatics, Novartis Institutes for BioMedical Research

Andrew Kasarskis, Ph.D., Scientific Director, Genetics, Merck Research Laboratories

Martin D. Leach, Ph.D., Executive Director Basic Research & Biomarker IT, Merck & Co. Inc. & Pfizer

Bruce Martin, VP Software, Complete Genomics

Kathryn Myers, Senior DBA, System Operations, National Center for Genome Resources

Jean-Christophe Olivo-Marin, Ph.D, Group Leader, Quantitative Image Analysis Unit, Institut Pasteur

Clifford Reid, Ph.D., Chairman, President and Chief Executive Officer, Complete Genomics

Alan Ruttenberg, Principal Scientist, Science Commons

Christopher Southan, Ph.D., EMBL European Bioinformatics Institute (EBI)

Dietrich Stephan, Ph.D., Co-founder and Chief Science Officer, Navigenics, Inc

Matthew Trunnell, Manager, Research Computing, Broad Institute of MIT and Harvard

Jonathan Usuka, Ph.D., M.B.A., Senior Director - Life Sciences Marketing, Accelrys

Antony Williams, Ph.D. Founder, ChemZoo, Inc.; Host, ChemSpider

Jeffrey Wisemen, Ph.D., CTO and Vice President, Discovery, Locus Pharmaceuticals

Matt Wood, Sequencing Informatics, Production Software, The Wellcome Trust Sanger Institute

For more speaker additions, please visit www.bio-itworldexpo.com

EVENT FEATURES:

- Access All Six Tracks for One Price
- Network with 1,500+ Attendees
- Hear 100+ Technology and Scientific Presentations
- Attend Bio-IT World's Best Practices Awards
- Connect with Attendees Using CHI's Intro-Net
- Participate in the Poster Competition
- See the Winners of the following 2009 Awards:
  - Benjamin Franklin
  - Best of Show
  - Best Practices
- View Novel Technologies and Solutions in the Expansive Exhibit Hall
- And Much More!

Register by January 23 and Save up to $500!

www.Bio-ITWorldExpo.com

Economic Development

A Celebration of Spanish Science

BIOSPAIN 2008 encouraged global biotechnology.

BY KEVIN DAVIES

GRANADA, SPAIN—In 2008, Spain celebrated stunning victories at Wimbledon, the Tour de France, and the European soccer championship. It appears the country's booming biotechnology sector has something to celebrate as well.

I'd tell you about it, but it did not occur to me prior to covering BIOSPAIN 2008—the country's premier biennial biotechnology congress—that virtually all the proceedings would be held in Spanish. (Curiously, the invitation from the organizers had omitted that snippet of information.)

Held in the heart of Andalusia in southern Spain, BIOSPAIN 2008 attracted 1,500 scientists and industry executives, with representatives of 162 companies and funding agencies partnering enthusiastically over tapas in the congress hall, while others engaged in nicotine networking outside the Palacio de Congresos. Attendance was triple that of BIOSPAIN 2006, a strong sign of the emerging state of Spanish biotechnology.

The largest crowd attended the formal inauguration of the conference, presided over by no fewer than nine government and local officials and foundation dignitaries, the main draw undoubtedly Spain's Minister for Science and Innovation, Cristina Garmendia.

According to the biennial report from Genoma Espana, one of the several agencies organizing the meeting, Spanish biotechnology employed 45,000 people in 2005 and projects that figure to surpass 100,000 by the end of this year. Turnover was 5.4 billion Euros, or 0.6 percent of Spain's GDP. Communications director Belen Gilarranz explains that Genoma Espana is a public agency established in 2002 that funds Spanish networks in genomics, proteomics, bioinformatics, and a DNA biobank. "We are the nexus between research and the market," says Gilarranz.

Two Americans provided the plenary entertainment at BIOSPAIN this year. Kenneth Morse, director of the MIT entrepreneur program, urged his Spanish audience to give voice to their entrepreneurial spirit rather than "retire or die—whichever comes first" in some large bureaucratic organization.

Asked how he balances work and family life, Morse said that it was quality time that mattered. Most families will welcome the individual reaching for their dreams, he said, before adding that that it helped if the spouse had a steady job. The Spanish government could do a lot for Spain, he said, by not changing the goalposts and of course providing additional funding.

Cristina Garmendia, Spain's Minister for Science and Innovation, was the center of attention at BIOSPAIN 2008.

Go Global

The other keynoter was life sciences investment banker Steven Burrill, who flashed through 150 slides in 50 minutes on the global state of health care and biotech. Noting that he helped write the business plan for Genentech, and advised the founders of Apple and Intel, Burrill echoed Morse's sentiments to "go for it." He emphasized the global world of biotechnology, and urged the audience to attack global problems, listing the transformation of medicine and health care he expects to see by the year 2020.

The 150 booths at BIOSPAIN 2008 featured some big names (Amgen, Roche, Thermo Fisher, Merck) and many smaller Spanish ventures. Biomol Informatics is a software consulting firm founded by Madrid scientist Paulino Gomez Puertas. His group offers consulting services in drug design and modeling, but he hopes to release packaged software next year for lead optimization.

Intelligent Pharma is a 13-person computational solution provider in Barcelona that offers services as well as software products: Helios is a ligand-based virtual screening program that relies on the compound's electrostatic potential. Selene is its receptor-based counterpart. Both work on the Daedalus grid computing platform, making use of unused CPU time.

Software company Integromics was co-founded by ex-IBM physicist Jose-Maria Carazo, a native of Granada. It is perhaps best known for its partnership with Applied Biosystems, but has just released a new product called Integromics Biomarker Discovery (IBD) for gene expression data analysis, developed with Spotfire, which performs statistical analysis of genomic data.

One of the current success stories in Spanish biotechnology is Madrid-based Zeltia, which has various subsidiary groups, including Genomica, a molecular diagnostics firm. Another is Pharma Mar, which discovers therapeutic compounds from marine organisms. Its lead cancer product is the recently approved Yondelis, a synthetic version of a compound extracted from a sea squirt, which is showing promise against ovarian cancer.

Merck is one of several global pharmas with a presence in Spain, but it made news recently by closing its Madrid headquarters. Merck is not abandoning Spain, but building a new center in Granada called Medina, in the region's thriving technology park. •

Dutch Drug Development Heats Up

Government, academia and industry stoke the life sciences fires.

BY ALLISON PROFFITT

The High Tech Campus at Eindhoven was started by Philips and now houses 51 companies.

THE HAGUE, THE NETHERLANDS—"Smoke from the Dutch chimney." That's what Willem de Laat, managing director of Top Institute Pharma (TI Pharma), calls the goal of life sciences research in The Netherlands, with a fire fed by the government, academia, and industry.

De Laat's organization is one of the big three public-private partnerships stoking the blaze. TI Pharma, The Center for Translational Molecular Medicine (CTMM), and BioMedical Materials Program (BMM), covering drugs, diagnosis, and devices respectively, are each set up with a funding model drawing cash and in kind resources from Dutch universities, the Dutch government including the Dutch Heart Foundation, Diabetes Foundation, and Kidney Fund, and private companies. Each was created and funded for five years initially, and projects depend on an internationally peer-reviewed call for proposals.

TI Pharma has funding of about $334.5 million for its 4-5 year commission and a directive to conduct groundbreaking, cross-disciplinary research in translational medicine, target finding, and biomarker validation, as well as improving the efficiency of drug development. Research at TI Pharma draws from a matrix of therapeutic areas and enabling technologies including autoimmune diseases, cancer, infectious, brain, and cardiovascular disease; and *in silico* modeling, biomarkers and bio-sensing. Each of the 42 projects launched in 2007 must have three partners including both academia and industry. The impressive list of companies on TI Pharma's partner list includes GlaxoSmithKline, Lilly, Amgen, AstraZeneca, and Galapagos.

CTMM's focus is on early detection *in vitro* and *in vivo* and personalized treatments. With five-year funding of about $514 million, CTMM released its first call for proposals last fall and chose nine projects in cardiovascular, oncology, and neurology. The organization is based on the High Tech Campus in Eindhoven, and some of their fellow residents count among their partners: Philips, Organon, DSM, AstraZeneca and a bevy of smaller biotech companies.

BMM is focusing on drug delivery, organ replacement, and passive and active scaffolds in four disease areas: cardiovascular, musculoskeletal, oncology, and nephrology. With almost $116 million for the first five years, seven projects were selected from the first call for proposals, and BMM has attracted partners including Philips, FujiFilm, Medtronic, and several Dutch universities.

Genomic Infrastructure

While CTMM, BMM, and TI Pharma make up the translational research leg of the Dutch life sciences knowledge infrastructure, The Netherlands Genomics Initiative, founded in 2002, forms the exploratory research leg. Sixteen genomics centers make up NGI in the 2008-2012 funding cycle with a budget of $348 million. Centers are dedicated to medical systems biology, bioinformatics, proteomics, forensics, toxicogenomics, celiac disease, ecogenomics, and nutrigenomics. NGI also runs Centre for Society and Genomics.

The applied science leg of research resides in The Netherlands' science parks. Leiden Bio Science Park hosts approximately 60 life sciences companies, eight institutes, and Leiden University Medical Center. The work done there ranges from research and discovery, product and process development to clinical trials, manufacturing, and sales.

Several Dutch public research institutes are based in Leiden including TI Pharma. LifeLines is a public Biobank studying aging across 165,000 participants from three generations with data collection dating back 30 years. TNO is The Netherlands Organization for Applied Research, an independent research organization focusing on telecom, defense, life sciences, and other areas of science. TNO Pharma determines safety and efficacy of drugs and translates *in vitro* and preclinical *in vivo* findings to humans.

The High Tech Campus at Eindhoven was started by Royal Philips Electronics in 1999. Today the campus is home to 51 residents including Philips, Sun, Accenture, and IBM, along with independent research organizations including CTMM, BMM, and the Holst Centre, an R&D institute that develops generic technologies for autonomous wireless transducer solutions. The ambition is to become a leading hotspot for high tech human focused innovation.

Maastricht sits at the heart of the Meuse Rhine Triangle, an area covering German, The Netherlands, and Belgium. The area is home to 23 medical device companies including Medtronic, and 48 other life sciences companies including genomic and proteomic research, plant biotech, nutraceuticals, and regenerative medicine research. •

A Scandinavian Bridge to Somewhere

Denmark and Sweden focus on personalized medicine and innovative science.

BY ALLISON PROFFITT

COPENHAGEN—Back in the 15th century, Denmark and Sweden were both part of the Kalmar Union. But that ended badly. The rulers in Denmark couldn't control the Swedish nobility and a bloodbath ensued.

Today, however, things are going much more smoothly for the two countries. Medicon Valley, a biotech cluster in the heart of the Øresund Region comprising greater Copenhagen and the surrounding region in Denmark and Region Skåne and Malmö in Sweden, boasts five science parks, 26 hospitals, 5,000 life sciences researchers, 41,000 life sciences employees, 14 universities, and 3.2 million people.

The area is "easy," advocates agree. National health care, work/life balance, progressive scale taxes, lean society, and lean businesses lead to a work culture that is deadline-driven and team oriented. Sixty percent of the life science companies in Scandinavia are based in the Øresund Region, and the impressive Øresund Bridge between Malmö, Sweden, and Copenhagen, Denmark, carries commuters back and forth to opportunities in both countries.

The Burrill European Life Sciences report last year dubbed Denmark the head of personalized medicine in Europe, and the country is home to Dako, the company along with Roche responsible for the HercepTest to identify breast cancer drug Herceptin's likely usefulness. It also boasts "embryo to grave" health records (mostly electronic), cheap and fast clinical trials, and a cancer register containing information on cancer and related diseases in Denmark since 1943 with compulsory disclosure for the past 20 years.

Venture capital in Denmark is less common than foundation funding. The Danish Cancer Society, a private membership society, spent more than $34 million on research in 2006, and finances 50 percent of the cancer research in Denmark. The Danish National Advanced Technology Foundation instituted a microRNA Research Consortium with Santaris Pharma and the University of Copenhagen to the tune of about $3.3 million.

The Øresund Bridge is 25,738 ft long, the longest combined road and rail bridge in Europe.

Vaekstfonden, a state-owned fund, facilitates financing in terms of start-up equity and high-risk loans for life sciences firms, and Denmark placed royalties from North Sea oil into a high-tech fund for local research. The Novo Nordisk Foundation Center for Protein Research was created with a donation of about $107 million strictly dedicated to protein research.

The country attracts some big names. Biogen Idec relocated a manufacturing plant to Denmark in 2004. The company looked at cost, R&D access, and the quality of culture and chose Denmark to house 3900 employees, 60% of which are Danish. Several recent Big Pharma and biotech deals have also featured Danish companies: Symphogen has entered a collaboration with Genentech; Nuevolution signed an R&D agreement with Merck; and Santaris Pharma signed a GSK deal.

Swede Science

The Swedes, on the other hand, are perfectly happy with their own accolades. It should be no surprise that the country that gave us furnishings giant IKEA has gained a reputation for innovative and functional design. It brings the same sensibility to science.

The country boasts several incubators and numerous startups. Young companies spun out of Lund University include Respiratorius, focusing on small airway treatments for COPD and asthma; SpectraCure, a device company specializing in photodynamic therapy for the treatment of tumors; and Exini, computer-aided diagnostics in the heart and brain.

IDEON, Scandinavia's first science park, hosts more than 250 companies, 3,000 employees, and 100,000 square meters of lab space. Life sciences account for about 30 percent of the tenants, with IT and telecom holding slightly more space. Ericsson started in IDEON in 1983; Bluetooth technology was born there. Sony Ericsson now employs more than 3,000 in the area. The slightly newer Medeon Science Park is limited to the life sciences. The park currently houses about 30 companies focusing on pharma and dental research. Nearby, Lund University supports three incubators (one at IDEON) and commercializes 30-40 projects annually.

With 30-40 startups waiting for incubator space and another 65,000 square meters of lab space under construction, IDEON hopes to double in size over the next ten years. And that sort of growth takes investment. Venture Cup and other competitions help raise capital along with grants, soft loans, pre-seed funds, angels (about 100 locally), and about ten substantial Swedish venture capitalists investing in life sciences. •

Bio·IT World
Signature Series

High-Performance Computing

Even in the coolest of economies, spending and investment can be justified and supported for productivity driving business solutions. So it goes for high-performance computing (HPC) technologies across a range of industries, and particularly the life sciences.

Today's biomedical and drug discovery researchers are often managing mountains of data that are increasingly growing from levels of terabytes to petabytes. Now, more than ever, they are turning to the developers of HPC solutions for manageable, expandable solutions that allow them to focus on turning data into discovery.

The HPC ecosystem covers servers to storage and the supportive and connective technology tissue in between, such as networking devices, accelerators, software, and a range of infrastructure options. This value chain, once properly enabled, can offer users substantial competitive efficiencies. Research firm IDC states that 97% of businesses they have surveyed using HPC technologies say they can no longer compete or survive without them.

IDC also indicates over the five year period from 2002 to 2007, the HPC market grew an aggregate 134% at an average compounded rate of 18.8%. In 2008, IDC predicts the HPC market will pass $12 billion heading toward $15 billion by 2011. Other HPC predictions include that:

- Clusters will make up over 70% of the market.
- The sub-$50K workgroup technical server segment is well poised for continuing growth.
- HPC users will increasingly look for relative ease of installation, operation, and upgradeability.
- Accelerators will aid progress in scaling application performance on multicore processors.
- Electrical power is fast becoming a central HPC issue, and from 2009 we will see more emphasis on power-efficient HPC systems.

- The explosion of data largely driven by technical computing will cause storage to grow even faster than HPC servers, reaching about $4.8 billion in 2008.
- Storage systems could well become more task based and focused on managing specific information categories (such as documents, database, technical drawings, Web), and therefore connect more directly to the source of the content.
- Online storage services will also proliferate with certain users looking to increase their storage efficiencies.

The benefits of HPC in science are well known. Many examples of improvements in research quality, collaboration, process acceleration, improved accuracy, and workflow efficiency have been demonstrated. This all leads to continued progress in enhancing productivity, and opening up more time for science.

In the following pages, we showcase some progressive HPC offerings through the eyes of the companies that have developed them, along with insights into ways users can benefit from their implementation. ■

Bio·IT World Signature Series High-Performance Computing

SGI

Itanium in Pharma and Higher Ed

Research conducted by pharmaceutical companies and in higher education has a common basis today.

In both realms, new lab equipment that produces vast amounts of data per experiment and rapidly growing public databases are contributing to a data explosion. The heart of most scientific research involves modeling, analyzing, and visualizing this data. And there's the commonality.

As the data volumes grow, so too does the complexity of the applications making use of that data. For example, researchers keep evolving their models and simulations to include finer spatial or temporal resolution or they extend the simulations to cover greater time scales. This helps bring the simulations closer to real-world conditions.

Speed of execution, which has always been an issue, takes on new importance today. In the past, the yardstick was how fast you could run one job. Now, the emphasis has shifted to workflow or how fast you can run a number of jobs where each is reliant on the output of the others.

What's needed to address these issues are high performance computing (HPC) solutions that combine the appropriate processing power with memory that can handle large datasets, storage whose performance matches the CPUs, and an infrastructure that acts as a high-speed data conduit and not a bottleneck.

And any discussion about HPC today must factor in energy. HPC systems have been notorious for consuming large amounts of electricity for power and cooling. Any new HPC deployment must opt for energy efficiency afforded by new processors and systems.

Matching applications and HPC resources

Increasingly more powerful computing resources must be applied to today's mix of more realistic and complex scientific research and development investigations. The challenge is how to accommodate the broad mix of applications that are typically used.

In particular, to ensure an application runs efficiently, there needs to be a matching up of an application's characteristics with suitable high performance computing (HPC) system components. For example, a life sciences application like BLAST or and a computational fluid dynamics application like FLUENT can benefit from a cluster approach where each node has sizable memory to support the large databases that analysis samples are normally run against. With this type of configuration, a large run can be broken into many smaller jobs all of which can be run simultaneously on discrete nodes.

In contrast, a Finite Element model of, for example, a climate model or a protein's interaction with a cell's membrane would require a system with a large share memory and tightly coupled nodes. Similarly, many Gaussian and HMMER runs can benefit greatly from share memory systems.

Essential Elements of Scientific Research HPC Systems

- **64-bit capability**
- **Support for large memory**
- **High performance switching**
- **Cost-effective storage systems**
- **Tight integration of all elements**
- **Software to enable workflows**

Essentially, an organization might devote a different computing platform to each type of application. To make the match, an application would be checked to see if it was memory-intensive, bandwidth-intensive, CPU-intensive, or a combination of all three. Other factors would also need to be taken into account. For example, how interdependent or loosely coupled are the nodes for a particular application? What volumes of data must be retrieved from storage to perform a calculation? How much output data is involved? Is the output data used in another application? Perhaps it needs to be combined with or compared to another database. Does the data need to be visualized?

Having determined an application's characteristics, it could then be assigned to a suitable computing platform. Obviously, it would be ideal if each application had its own perfectly optimized platform available when needed. Of course, that would be impractical. So what's needed is a computing platform that incorporates a variety of computing technologies. Such systems might include the use of multi-core processors, shared memory, discrete cluster nodes, and hardware accelerators.

Focus on the processor: Enter Itanium

One point in favor of using HPC clusters is that, increasingly, a variety of newer generation components have been incorporated into these clusters.

Given the large datasets researchers typically work with and the need for high processing power, many groups are opting for systems based on Intel Itanium processors. The true 64-bit nature of the Itanium allows support for very large amounts of memory. The advantage here is that loading data into memory rather than drawing it off of hard disk drives or storage systems significantly speeds the execution of many

High-Performance Computing

applications.

In practice, the true performance will depend on many factors, including the amount of memory a company is willing to purchase and whether applications are optimized to take advantage of the processor's capabilities. Additionally, if a 64-bit server is part of a cluster, the efficiency and performance of the interconnection software and switches in the cluster will also affect performance.

That said, the main advantage of using a 64-bit system is that much more data can be put into memory — which means faster results. A 32-bit system can access only 4 gigabytes of memory. In many common research science applications — searches of a huge database, for example — this 4 GB limit can be a real performance stopper.

To work with a large database or other big application requires the use of virtual memory, which sets aside a portion of a hard drive's space to temporarily store additional data for the application. But accessing and manipulating data stored in hard drive virtual memory in this way takes about 40 times longer than when the data are stored in random access memory (RAM). A 64-bit system circumvents this obstacle because it can accommodate a much larger amount of data in memory — up to 16 terabytes. By loading an entire database into memory, any calculations, searches, and manipulations can be performed more quickly than when accessing the same data residing in virtual memory on a hard disk drive.

Three general classes of applications benefit from the 64-bit processor:

- Widely used search and sorting bioinformatics algorithms such as BLAST, Hidden Markov Modeling (HMM), and Smith Waterman
- Simulation and modeling applications such as molecular docking and computational chemistry algorithms, as well as programs that calculate the binding energy between two molecules
- Applications that include large-scale data mining (and searching), knowledge management, and visualization.

Multi-core moves ahead

Many cluster nodes today use multi-core processors to increase the raw processing power without increasing the server size, which is becoming an issue in crowded data centers. Multi-core processors often run at lower clock speeds than their comparable single-core processors and as a result, they often consume less electricity, which is also a growing consideration in many data centers. (And as a result of using less electricity, the cooling demands are also lower, thus saving additional energy.)

The true 64-bit nature of the Itanium allows support for very large amounts of memory.

Multi-core processors also typically come with integrated memory controllers to improve the flow of data between a CPU and its associated memory. This can help improve the performance of many classes of applications by eliminating the wait for data to be fed into today's high performance processors.

One factor limiting the most efficient use of multi-core processors today has nothing to do with technology. It involves software licensing. Some vendors license their software by the number of processors the application is running on; others by the number of instances of the application. If a multi-core system is being used in conjunction with virtualization, this might complicate matters.

Naturally, this is not an issue with the open source software widely used in scientific research. But it can become a problem with commercial applications. There may be cores available, but no licenses to run the application.

In many cases, even as CPU performance increased, some science applications could use an additional boost from co-processors. In the past, some applications were accelerated using application-specific integrated circuits (ASICs) and field-programmable gate arrays (FGPAs). And many industry experts believe researchers will soon be tapping the power of graphics processing units (GPUs), which due to the demand for more realistic gaming products, have been growing in power at a faster rate than traditional CPUs.

These approaches have the potential to deliver additional benefits. For example, as is the case with multi-core processors, FPGA systems are often very energy efficient on a performance/watt basis when compared to running the same job on a traditional server.

The pure increase in processing performance new CPUs and accelerators deliver is only one aspect that impacts the ability to run applications in a timely manner. A cluster's interconnection infrastructure must also be considered. This includes addressing the switching fabric and the performance of the storage systems that feed the data to life sciences applications.

For years, InfiniBand and proprietary interconnection technologies were used to reduce latency and increase overall performance. Increasingly, more systems today are incorporating standards-based interconnection technology that is based on gigabit Ethernet.

When choosing an interconnection technology, organizations try to find the best match performance-wise, but other factors may also come into play. For instance, IT staffing skills are often considered (many more IT people have experience with Ethernet than other technologies). Similar considerations are often made today for matching storage systems with applications. Namely, the systems are selected based on performance criteria and staffing expertise criteria.

And for some applications, shared memory systems offer additional benefits to increase the performance of an application.

New drivers emerge

The data explosion is causing changes in another area: Storage.

All of the data must be stored somewhere. And even though the price per TB of storage has dropped significantly over the last few years, other factors must be considered when determining what type of

systems will be used to hold the data.

One factor, noted above, is matching storage system performance to an application's requirements. Is the data needed for modeling, analysis, or visualization? If so, is that work part of a computational workflow? Such questions will help determine if data is stored on a high-performance storage system or a more cost-effective system that has lower performance.

A more recent storage performance consideration is how to accommodate the growing adoption of virtualization. With virtualization, multiple applications share a server's resources by running as virtual machines. Many organizations are consolidating applications that previously ran on discrete physical servers onto a single server. It is quite common for there to be a 10-to-1 consolidation.

The impact on storage systems is that when a server ran a single application, only that application was placing demands on the storage system for data reads and writes. If ten applications are running on each server node in a cluster, the I/O demands on the storage system can be increased significantly.

Beyond storage and virtualization, the other relatively new driver with HPC systems is workflow.

Today what is important is not how fast a single job can run. The objective is to maximize the number of specific jobs, related to one research endeavor that can be run in a given time. In other words, the goal is how fast a researcher can reach a decision about what to do next based on running a series of computational and visualization applications.

Platform variety is the key

What's needed to accommodate the varying demands for HPC resources is a new platform that supports all types of HPC hardware systems including shared memory, multi-processor clusters, and accelerators like FPGAs.

The hardware itself must include software that helps manage and automate the process of running jobs. The idea here is to take the decision making process that matches an application and specific job to particular computing resources out of the hands of the scientist and make the process automatic based on some prior knowledge and rules.

With these points in mind, SGI has developed the Hybrid Computing Environment. The SGI Hybrid Solution system incorporates SGI's shared memory Altix, HPC cluster Altix XE, and FPGA-based RASC.

The system uses Moab as its intelligent interface software. Moab can schedule jobs to run on multiple architectures and uses policies to automatically determine the best match of an application to an HPC architecture.

As data volumes grow, so too does the complexity of the applications making use of that data.

To run jobs using Moab, users access a Web portal. Managers can create custom templates for common jobs where users do not need to know the characteristics of an application and which HPC resource is well suited for it.

Moab allows researchers to run a variety of applications types from a single interface in a more efficient manner. In particular, it can accommodate jobs traditional run in batch mode. However, Moab can consider workloads and resource usage before running the jobs and schedule their running in a way that optimizes throughout.

Additionally, Moab can be used to allocate appropriate resources for interactive jobs such as visualization applications where the researcher must examine a result to initiate the next computational or visualization step. This enables visualization to be brought into the same cloud of HPC resources as other applications. This is as opposed to having a dedicated system just for visualization as is often done. The benefit here is that it makes more efficient use of HPC resources and perhaps has the potential to expand the use of visualization as a dedicated system might not be cost effective on its own.

Benefits of the Hybrid Computing Environment

There are a number benefits to using the SGI Hybrid Computing Environment.

First, organizations can improve their computational workflows. Applications can be assigned resources and scheduled to run based on computational demands. This can significantly reduce the total time to run all jobs when a mix of applications is used.

Second, the SGI Hybrid Solution optimizes the use of HPC resources. As noted, batch and interactive jobs such as visualization applications can share the same computing resources in contrast to having separate systems for each. Additionally, applications that benefit from shared memory, FPGAs, or other technologies can automatically be allocated based on HPC requirements and schedule to run in a sequence that optimizes use of the available resources.

A third benefit is that the SGI Hybrid Solution lowers total cost of ownership. There are fewer discrete systems to manage, so the burden on IT staff is reduced. Additionally, the variety of HPC resources are all managed using a single console using similar commands.

Fourth, user training is significantly reduced. Researchers and engineers do not need to develop an expertise in computer science to run their applications in the most efficient and timely manner. Managers and IT staff can set policies for matching applications to resources, scheduling jobs, and other tasks. The researcher or engineer simply uses a Web portal to submit a job and the system takes it from there.

Since the SGI Hybrid Solution concentrates HPC resources under one system, it allows all those resources to share common data. This means an organization can vastly simplify its data infrastructure by storing all data in one place. This also simplifies administration of data storage devices and the management of tying data volumes to specific applications.

And finally, by virtue of all of these other aspects the SGI Hybrid Solution helps improve research and engineering workflows.

High-Performance Computing

Jobs are executed in a timelier manner making the best use of HPC resources. The real-world impact is that complicated and multi-step analysis routines, as well as interactive visualization jobs, are completed in shorter times helping researchers make more intelligent and faster decisions as to how to proceed.

All of these benefits can add up to a significant advantage by accelerating work on today's complex scientific research problems. ■

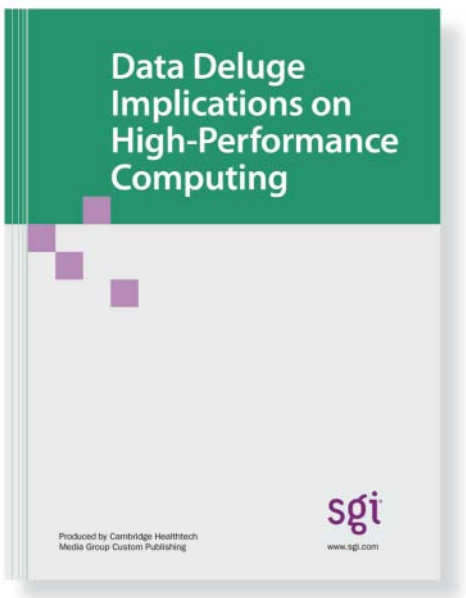
sgi® INNOVATION FOR RESULTS™

intel Itanium inside™

Available soon from SGI
A new whitepaper

Data Deluge Implications on High-Performance Computing

The explosive growth in data associated with life sciences and academic scientific research is placing new demands on high performance computing systems. Simply leveraging the ever-growing raw processing power of new CPUs is no longer sufficient. Increasingly, HPC systems must tightly integrate cost-effective, high- performance storage systems to ensure optimal analysis, modeling, and visualization workflows.



Reserve your complimentary copy now by requesting it from:
Lisa Provenzano-SGI
lprov@sgi.com

For more information on SGI's High-Performance Computing solutions visit www.sgi.com

Bio·IT World Signature Series
High-Performance Computing

MITRIONICS

Innovating Software Acceleration for Genome Informatics

How would you describe the current state of the industry for accelerated computing technologies?

When Intel and AMD opened up their architectures to support acceleration technologies such as FPGAs and GPUs, it marked the beginning of a new era called "hybrid computing," where computing systems support multiple types of acceleration technologies and processors, allowing each to excel and be applied to the application areas best suited to its architecture. Accelerated software applications typically perform 10x to 100x faster than their non-accelerated versions. And acceleration technology fits nicely with the "green computing" movement by saving up to 90% of electrical consumption compared to a comparable cluster providing the same performance benefits.

Because of Intel and AMD, and the adoption of their standards by major OEMs, researchers, developers, and ISVs can now purchase commercial off-the-shelf hybrid computing servers and development systems from leading vendors that include integrated acceleration technology. And these new systems are available at significantly lower costs than the previous generation of accelerated computing systems and solutions.

Which acceleration technologies are best suited to life sciences and bioinformatics applications?

When interfaced closely with multicore systems, FPGAs have an eight to ten-year track record of proven success in life sciences. This gives them a head start in terms of identifying the target applications and algorithms to be accelerated, and with the availability of parallel programming languages used to accelerated software.

However, GPUs and other technologies are generating interest on the PCIe Bus, particularly in molecular dynamics. This is very healthy for the overall hybrid computing environment.

Name some recent industry developments that have changed the accelerated computing industry?

Several companies have developed products featuring new form factors for integrating FPGAs directly within the system's motherboard for low latency, high bandwidth and high performance while generating very low heat. This is the combination needed for true hybridization. The newest generations of PCIe cards have much higher bandwidth as well. Mitrionics is working closely as a partner with all of the major players and smaller companies in this arena.

Mike Calise, Executive Vice President and General Manager for Mitrionics, Inc.

Nallatech is well recognized in the industry for their FPGA-equipped PCI and PCI Express cards. And XtremeData is now making a big impact with their FPGA-enabled Front Side Bus (FSB) modules that plug directly into the motherboard of systems. Both of these solutions enable FPGAs to be integrated into industry standard systems at a lower price point than previously possible. Most importantly, these technologies are backed by Altera and Xilinx who know a bit about driving FPGA densities and will continue to do so.

What has Mitrionics done to take advantage of the current technology evolution and where do you see yourself in the mix of things?

Our recent introduction of "Hybrid Computing Development Systems" based on servers from HP and Intel is designed to drive innovation in life sciences as well as other industries. It provides a complete software acceleration development environment for ISVs, researchers, and developers with price points starting at $15k.

The systems are equipped with tightly-integrated Mitrion Virtual Processor, (MVP) with FPGA modules from XtremeData and include the complete Mitrion Software Acceleration Platform with Mitrion Language, the Mitrion SDK, sample code, and algorithms.

What are the business and technology segments within life sciences and bioinformatics that have the most explosive growth potential?

In the near future, preventative medicine and personal genomics will begin to make a major impact on society and medicine. The fields of healthcare and genomic research are converging and will soon be applying therapeutic strategies to improve health through preventative medicine. Next generation sequencers will be a key component in the future success and widespread adoption of this new discipline and are ideally suited to benefit from acceleration technology. Mitrionics sees this as an area of strong interest for our company and the application of our technology. Without our technology offering, the data crunching capabilities of NGS equipment could fail to reach its full potential, or even stall. We can help unlock the compute side of the equation. ■

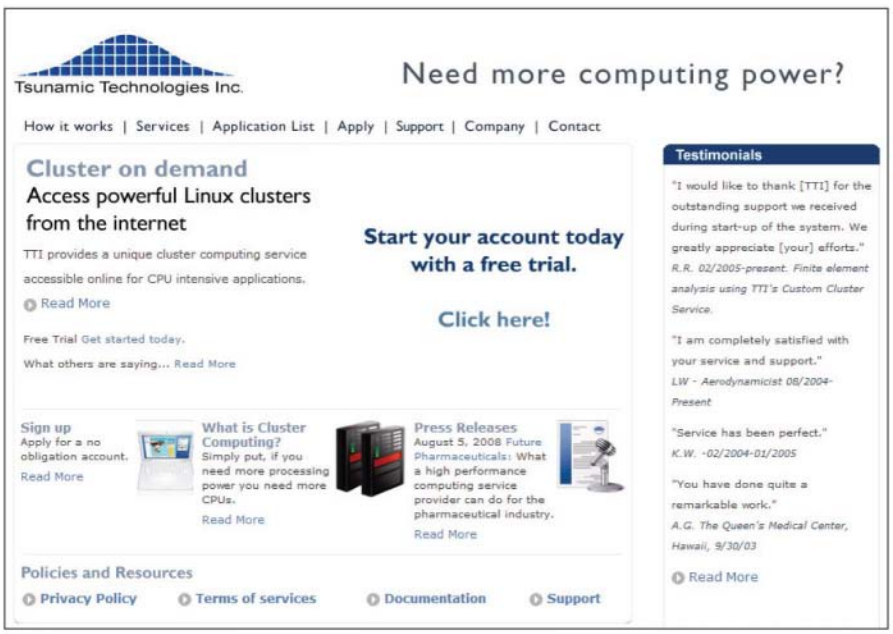
TSUNAMIC TECHNOLOGIES INC.

High Performance Computing on Demand

What services does TTI provide and how do your customers uses these services?
Tsunamic Technologies Inc. (TTI) is a leading provider of High Performance Computing services for small to medium sized organizations. Our Life Science customers use our services in much the same way as they traditionally would with an in-house Linux cluster solution, but TTI allows them to forgo the costly task of purchasing and managing a HPC facility themselves.

What applications and tools are available on your HPC clusters that are of interest to the Life Sciences?
TTI can install virtually any software whether it be Open-Source, Commercial, or internally developed. Some of the popular life science applications used by our customers include Blast, mpiBlast, HMMR, Abinit, Gromacs, AutoDock, ClastalW, and InterProScan. We also provide a wide variety of programming languages and optimized scientific libraries such as MPIch, OpenMPI, LAM, ATLAS, LAPACK, BLAS, and NetCDF. There are many others and the list is always growing.

What are the advantages of using a Compute-On-Demand service rather than an in-house solution?
Using our High Performance Computing service has many benefits over purchasing or leasing hardware. Foremost, initial costs are reduced dramatically. By treating CPU-Hours as a consumable commodity, budgeting and planning become much easier and more flexible. A company not only gets access to the hardware it also gains the experience and knowledge-base of its HPC service provider. TTI has been able to suggest to its customers more advanced and technically challenging solutions to their unique problems, thereby saving them time and money. Another benefit to using an HPC service provider instead of spending the extra effort to house your own cluster is that you eliminate the risk of locking into a single architecture. By using an HPC provider such as TTI, you reduce your risk, have access to it's knowledge-base and experience, and save time and money.



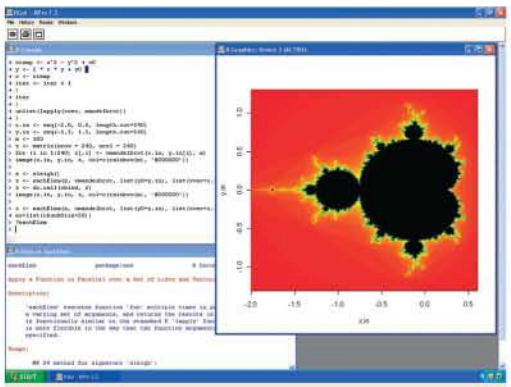
Sign up for an open trial account at our website.

What do you see as the future for HPC in the Life Sciences and what is TTI doing now to prepare for this?
As the Life Sciences advance their technologies and expand their offerings, their storage and computational requirements will indubitably increase tremendously. In order to keep pace in a competitive market, they will need faster turn-around times for the R&D and production computational projects. Thankfully, the HPC industry is continuing its development of new technologies to meet these demands including GPUs, Cell Processors, and ultra fast storage networks. These technologies are well suited for scientific and structural engineering applications; ideal for life sciences, biotech, and pharma industries. TTI can stay ahead of the technology curve and offer access to these new technologies at a cost below that of a similar in-house solution without any investment risk to the customer. ■

Tsunamic Technologies Inc.

WWW.CLUSTERONDEMAND.COM

New Products



Open Source Power

REvolution Computing has launched ParallelR version 1.2. ParallelR enables users of the open source R program to complete statistical analyses much more quickly by taking advantage of multiprocessor systems and network systems. Computations can be run across heterogeneous hardware configurations including 32 bit, 64 bit, SMP, and cross-platform configurations such as combinations of Linux, Windows, and Mac OS X operating systems.

Company: REvolution Computing
Product: ParallelR
Available: Now
For More Information: www.revolution-computing.com

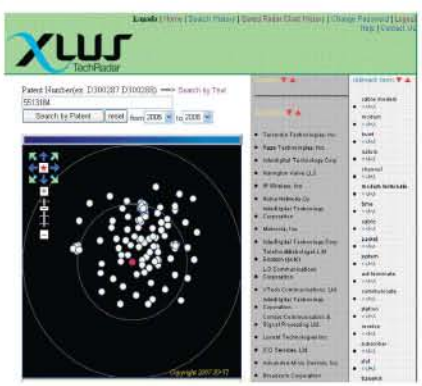
Parallel Analysis

Parallel Line Analysis has been incorporated into Gen5 Data Analysis Software version 1.06, a BioTek Instruments product. Parallelism analysis can be made with Gen5's linear regression, 4-parameter and 5-parameter logistic fits. Multiple curve metrics and detailed statistics are available to ensure the most accurate estimations of parallelism and relative potency between the experimental curves and control curves. Along with parallel line analysis, the Gen5 update offers a number of enhancements including color gradient data views, individual patient reports, use of curve parameters in data transformations and more.

Company: BioTek
Product: Gen5 v1.06
Available: Now
For More Information: www.biotek.com

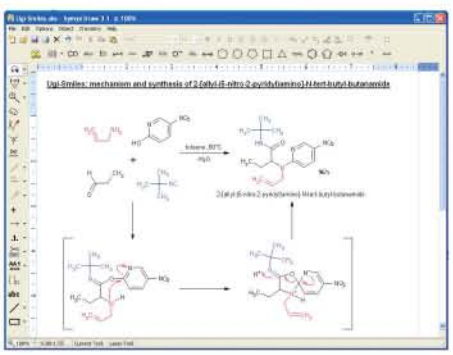
Patent Visualization

InfiniteBio has introduced XLUS Eagle, a web-based software tool that can visualize massive numbers of patent searches and analysis results. To illustrate the similarity and technological relationships between patents, XLUS produces a radar chart in addition to an ordinal patent list. Core technologies include natural language processing, data informatics, visualization and easy expression, and IP knowledge and analysis.



Company: InfiniteBio
Product: XLUS Eagle
Available: Now
For More Information: http://so-ti.com/en/products/xlus/

Structure Drawing Updates



Symyx has announced Symyx Draw 3.1, software for drawing chemical structures and searching, registration, and collaboration. Draw 3.1 is a replacement for ISIS/Draw and while it looks and feels the same, Draw 3.1 has an easier to use drawing tool, toolbar, and library of protecting group templates. Draw 3.1 also features the ability to utilize built-in structure-to-name and name-to-structure converters, create and edit polymers, mixtures, and formulations, create and edit Markush (Rgroup) queries, and build and visualize 3-D chemical structure queries.

Company: Symyx
Product: Draw 3.1
Available: Now
For More Information: www.symyx.com

Array and Sequencing Analysis

Illumina has launched its GenomeStudio data analysis software designed to analyze both microarray and sequencing data. GenomeStudio enables the correlation of biological variation across multiple applications and graphical display of the results. GenomeStudio also allows researchers to integrate with third-party software providers for advanced downstream analysis. In addition to algorithms for copy number variation detection, SNP calling, and differential analysis, GenomeStudio contains tools for visualizing sequencing data that are fueled by a new and expanded version of Illumina's Genome Analyzer Analysis Pipeline.

Company: Illumina
Product: GenomeStudio data analysis software
Available: Now
For More Information: www.illumina.com

Educational Opportunities

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

Featured Events

Supercomputing, Nov 15–21, 2008 Austin, TX
Supercomputing is the premier international conference covering high performance computing, networking, storage, and analysis. Attend this annual conference for technical and educational programs, workshops, tutorials, an exhibit area, demonstrations, and hands-on learning. Visit *http://sc08.supercomputing.org/*

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

The 4th Annual World Healthcare Innovation and Technology Congress, Dec 8–10, 2008, Washington, DC
This annual conference—WHIT v.4.0—is a must-attend event focusing on innovation and technology for healthcare C-suite executives. Visit *www.whitcongress.com*

LabAutomation 2009, Jan 24–28, 2009, Palm Springs, CA
The Association for Laboratory Automation (ALA) presents the 2009 annual LabAutomation event. It is the world's largest conference and exhibition focusing exclusively on the rapidly growing field of laboratory automation. A diverse group of academicians, scientists, engineers, business leaders, post-docs and graduate students from around the globe attend this annual conference. Visit *www.labautomation.org/LA09*

TEPR 2009, February 1–5, 2009, Palm Springs, CA
TEPR is a premier health IT conference for healthcare and IT professionals seeing innovative and practical solutions to current health IT issues. Topics will include, the role of cell phones in healthcare, linking physicians to hospitals and the patient centered medical home, and other critical topics related to health IT. Visit *www.TEPR.com*

The World Health Care Congress 2nd Annual Leadership Summit on Consumer Connectivity & Web Empowerment, February 23–24, 2009, Carlsbad, CA
This summit offers compelling strategies for today's providers and insurers to revolutionize healthcare through the integration and adoption of today's emerging and, at time, disruptive technologies. Visit *www.worldcongress.com/connectivity*

CHI Events

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

High Content Analysis
January 5–9, 2009 • San Francisco, CA

PEPTalk 2009
January 11–16, 2009 • Coronado, CA

Next Generation Antibodies
January 14–16, 2009 • Coronado, CA

Biomarker Assay Development
January 26–28, 2009 • San Diego, CA

ASAP Alliance Summit
February 9–12, 2009 • Fort Lauderdale, FL

Molecular Medicine Tri-Conference
February 24–27, 2009 • San Francisco, CA

Stem Cells Congress
February 25–27, 2009 • San Francisco, CA

Translational Medicine
February 25–27, 2009 • San Francisco, CA

Drug Development Latin America
March 9–10, 2009 • Miami, FL

Next Generation Sequencing
March 17–19, 2009 • San Diego, CA

Barnett Educational Services

For more information on Barnett live seminars, interactive web seminars, on-site training programs, and customized eLearning development services, please visit **www.barnettinternational.com**

Live Barnett Seminars

Drug Approval Process
December 2–3, 2008 • Chicago, IL

Conducting Clinical Trials Under ICH GCP
December 4–5, 2008 • San Diego, CA

Working with CROs
December 9–10, 2008 • Boston, MA

Conducting Clinical Trials in Emerging Regions
December 11–12, 2008 • Boston, MA

Drug Development and FDA Regulations
December 11–12, 2008 • Boston, MA

International Drug Approval Regulatory Procedures
December 18–19, 2008 • Philadelphia, PA

PharmacoKinetics
February 26–27, 2009 • Philadelphia, PA

Interactive Web Seminars

Meeting International Safety Reporting Requirements
November 25, 2008

Winning Strategies for Clinical Recruitment and Media
December 2, 2008

Electronic Medical Records and Source Document Verification
December 9, 2008

The Role of Business Process Management in an FDA Regulated Industry
December 10, 2008

Comparing FDA and Health Canada Regulations
December 16, 2008

Examining the Impact of the eCTD on the Regulatory Submissions Process
December 18, 2008

Life Science Jobs

Memorial Sloan-Kettering Cancer Center – Linux Systems Administrator
MSKCC Bioinformatics Core in Manhattan seeks Linux System Administrator. Install, configure, update, monitor and troubleshoot HPC cluster, email and web servers. Master's degree and 5 years experience with HPC, network security, Postfix, Apache, Tomcat, Grid Engine, Solaris. Email *bicjobs@cbio.mskcc.org* or *www.mskcc.org #015429*. EOE/AA

MedImmune INC. – Associate Director, Development Information Systems
(Gaithersburg, MD) Position is responsible for IS strategy for the global development organization, implementation and oversight of the ongoing projects and support of existing systems. If you are interested in this position, apply online at *www.medimmune.appone.com* and search for Req 01124.

White Papers, Webcasts, and Podcasts

Keep abreast of current industry trends and developments. Browse our extensive list of free Life Science white papers, webcasts, and podcasts on **www.bio-itworld.com**. Interested in learning about developing a multi-media solution to list in the magazine and on our web sites? Please email Sales_chmg@chimediagroup.com

White Papers

Storage for Science – Methods for Managing Large and Rapidly Growing Data Stores in Life Science Research Environments

Sponsored by Isilon

Large and rapidly growing stores of file-based and other data are a hallmark of life sciences research and bioinformatics. Determining how best to manage those data stores has become a significant challenge for researchers and IT pros alike. White paper covers: guidance on the many storage requirements common to life science research; an explanation of the evolution of modern storage architectures; and summarizes the major data storage architectures currently in use.

Addressing Life Sciences' Constantly Growing Data Challenges Research Environments

Sponsored by BlueArc

The continued explosion of raw experimental data, the increased use of video, the growing adoption of new data retention practices, and the move to high throughput computational workflows are all placing new demands on the way life sciences organizations store and manage their data.

Managed Innovation, Assured Compliance

Sponsored by SAS

Discovery organizations are identifying a lot of promising compounds, but clinical research processes haven't kept pace with timely testing of all those potential therapies. This white paper describes how SAS Drug Development supports true innovation across the clinical trial process.

Webcasts

Storage for Science – Methods for Managing Large and Rapidly Growing Data Stores in Life Science Research Environments

Sponsored by Isilon

Webcast covers: Guidance on the many storage requirements common to Life Science research; The evolution of modern data storage architectures; Major data storage architectures currently in use; and Isilon's IQ clustered storage product as a strong and flexible solution to those needs.

Podcasts

Hospital Paperwork No Longer has to be an In-Patient Procedure

Sponsored by Adobe

Paperwork redundancy has become a significant bottleneck in the healthcare system. This podcast covers how Adobe solutions for healthcare can help you streamline your paperwork and stop making paperwork an in-patient procedure.

Enabling Translational Research Informatics

Sponsored by GenoLogics

What are the challenges facing life sciences research labs today in terms of managing their data and facilitating their translational research vision? What are the trends we are seeing for organizations to adopt informatics solutions to solve these challenges and accelerate their research outcomes? Access this podcast to learn how life sciences research groups are successfully adopting a common informatics solution to enable their systems biology and translational research initiatives.

The Russell Transcript

Microsoft: Eyes on the LS Enterprise

JOHN RUSSELL

Many folks wonder what Microsoft's eventual play in the life sciences will be. Clearly biomedical research looks more and more like an exercise in digital content integration, management, and mining. Within the last few years, Microsoft—and others—awakened to opportunities in health care writ large and to biomedical research as a key piece of that enterprise. To a large extent, Microsoft's life science push has followed its traditional dependence in independent software vendor partners building solutions on top of Microsoft products. There is even a relatively new organization, the Microsoft BioIT Alliance, aimed at wooing ISVs and helping sell to life sciences (see, "Microsoft's Move into Life Sciences," *Bio•IT World*, Nov. 2007).

Yet the size and complexity of the biopharmaceutical industry presents other opportunities and challenges. Many large companies want enterprise-scale solutions which Microsoft would dearly love to sell directly. In health care, for example, Microsoft's Amalga "product" is an enterprise-scale data integration solution sold directly to big hospitals. Indeed, Microsoft is building a specialized version of Amalga for the life sciences, which will be ready soon.

Then there is the army of software developers toiling inside biopharma which Microsoft would love to persuade to write apps on Microsoft products.

> "[We] talk about solutions for the business...we try and focus outside of IT."
>
> **Michael Naimoli,** Microsoft

Health Care Missionary

Spearheading Microsoft's missionary work to biopharma is mostly Michael Naimoli's job. He is director of the life sciences group, one of three groups—providers; plans; and life sciences—embedded inside Microsoft's U.S. Health and Life Sciences organization which is GM'd by Steve Aylward.

Says Naimoli, "Seven years ago there were approximately five people focused on health care worldwide at Microsoft," says Naimoli. "Now there are more than 700. My group is responsible primarily for the enterprise space (in biopharma and medical devices), so these are the largest customers in the U.S. The goal of my crew is not really to go and talk about products, but to go in and talk about solutions for the business—so get out of IT and have conversations with the business decision-makers around what are they trying to do and how can Microsoft help.

"We try and focus outside of IT, so the vice president of clinical development, discovery decision sciences, sales and marketing, medical affairs, manufacturing—key decision-makers in those arenas. Occasionally the CIO—we certainly talk to the CIO and his or her organization, but our conversation is about how they service their customers within the organization and what their customers are looking for."

Naimoli won't say much about the forthcoming Amalga-like product, but talks freely about efforts to win over internal developers. "[These are] developers that for long time have been writing applications on our competitors' platforms. They are often fans of open source software. We do provide standards-based solutions for the industry," he says.

"The way we've been dealing with that group has been to go in and demonstrate what our platform can do with respect to data visualization and workflow and collaboration. We have them look at Windows Presentation Foundation (WPF), for example, and we're working to have them think, "I'm going to write to WPF or I'm going to learn about WPF." Or we have them look at HPC Server 2008 and say, "Maybe high-performance computing on a Microsoft platform is something that we have to look at and consider writing applications to it." That's becoming a bigger part of my group's focus," he says.

Building a History

Microsoft is also working to build its lineup of powerful case histories. Two arrows currently in the quiver are its work with Eli Lilly and Novartis. With Lilly, it was the Orion project in which PerformancePoint Server and Project Server were deployed to provide data visualization and project milestone information enterprise-wide. A Novartis, he says, "We built the decision support environment. That's a solution that pulls together project data and all of the documentation around the project data into a single system."

While Naimoli group has limited interaction with the BioIT Alliance, "I can tell you our plan in the not-so-distant future is to provide them with code that we're working on—live code around data visualization."

It will be interesting to watch this next stage of Microsoft's LS effort.

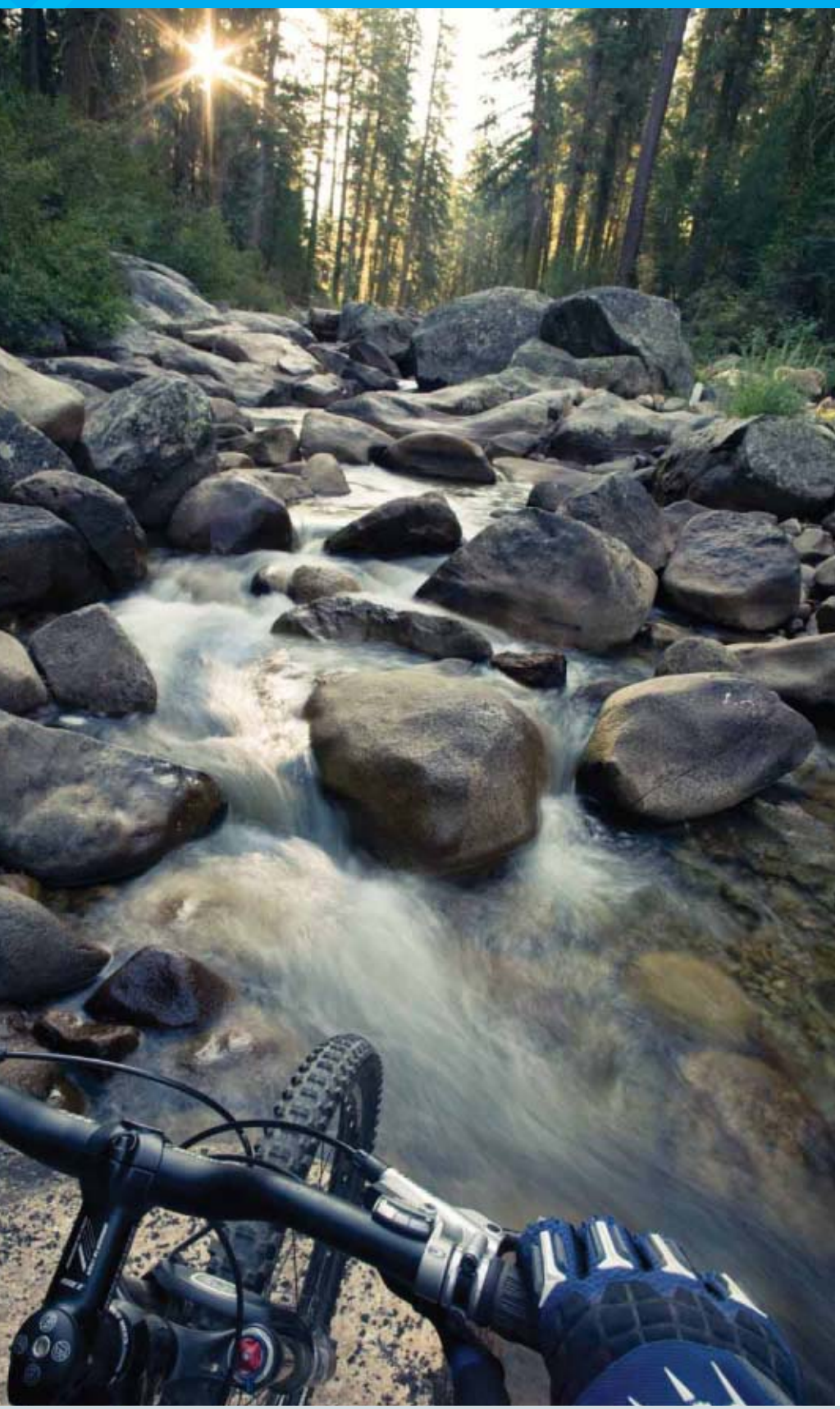
INTENSIVE CARE FOR CRITICAL DATA

With your data growing every day, the need for **reliable, scalable** storage solutions is critical. Our extensive range of high-performance solutions, including new Dell EqualLogic™ iSCSI SAN devices, deliver exceptional performance and functionality to simplify your data storage needs today and tomorrow.

SIMPLIFY YOUR DATA STORAGE AT
DELL.COM/Intensive or call 800.847.4156

DELL

Go beyond the everyday. EVERY DAY.



Symyx Notebook

The Freedom to Experiment.

Whether it's on the trail or in the lab, you want the freedom to take new approaches, routes, and paths to your goals. That's why there's Symyx Notebook. It's the only electronic laboratory notebook that can be deployed across the enterprise in multiple scientific disciplines. With Symyx Notebook, research teams share a single application to document, work, collaborate, and speed the experimentation workflow.

Symyx Notebook streamlines the capture of all experimental information and intellectual ideas. Everyday tasks such as data capture and note taking are optimized and automated. All of which gives you the time and freedom you need to experiment—and get back to doing science.

To learn more, visit www.symyx.com/notebook6

Symyx®