

## A second Human Genome Project?

With only 4% of the Human Genome Project completed, albeit ahead of schedule, one might conclude that the so-called genomics revolution has quite a bit more time on the meter. But there is another aspect of the human genome that we can no longer afford to ignore: The more than three million differences in the genetic code that uniquely identify each one of us. Not much when one considers a genome of over three billion bases, but of sufficient clinical importance to forever change the face of human healthcare and propel forward the field of pharmacogenomics.

Who should know better than Francis Collins, director of the US National Human Genome Research Institute, who recently helped Novartis initiate a summit meeting with their pharmaceutical brethren to discuss the joint funding of a single nucleotide polymorphism (SNP) mapping effort? According to Collins, a publicly available SNP map will be critical to understanding how to use the data that is generated by the Human Genome Project. A catalog of all common human sequence variations will enable the identification of weaker polygenic contributors to disease, the design of personalized prognostic strategies, and ultimately the adoption of individually optimized therapies—the Holy Grail of pharmacogenomic research.

In May, TIGR's Craig Venter declared his intent to mount a massive effort to sequence and collect SNPs, forming with Perkin-Elmer the \$200 million joint venture Celera Genomics. Venter has an acute awareness of the cutting edge in allying technology to science: If he is not the first to start something, he will do it better or faster. (In the case of Celera, he claims he can resequence the human genome ten times over and seven

years ahead of the public effort). The fact that Venter has put a high priority on mapping SNPs is significant.

And now Incyte has launched an initiative. For a mere "snip" (\$38 million), it has acquired Hexagen—a UK company with a proprietary high-throughput SNP scanning system—and intends to invest over the next two years a staggering \$200 million in an effort to sequence the protein-coding regions of the human genome and "gather SNP data for every human gene."

So it seems that Daniel Cohen and Genset got the concept right a year ago when they announced their effort to create a biallelic map of the entire human genome. Now the race is on to identify SNPs of clinical relevance—perhaps a trickier proposition than many of the companies involved would like us to believe, particularly as the technologies are unproven and the existing data preliminary. But with the US National Institute of General Medical Sciences promising to galvanize government-sponsored research in the area\* and an influential nucleus of private companies investing significant resources, there are interesting parallels with the early days of genomics. Back then, it took private and public money to kickstart the field. In the light of recent events, the move into SNP mapping and pharmacogenomics—the second Human Genome Project if you will—should be as swift and dramatic as the events that followed the "original" Human Genome Project.

\*Rochelle Long, chief of the Pharmacological & Physiological Sciences Branch, will present NIGMS' plans at Nature Biotechnology's Validating Pharmacogenomics meeting next month (October 17, 1998).

## Beat bioterror with batch science

The US government's multimillion dollar antibioteerrorism plan to stockpile vaccines and/or antibiotics at strategic locations around the country in order to protect civilian populations in the event of a bioweapons attack won't work. It is the equivalent of having the citizenry rush to their basements and throw their arms up over their heads during a nuclear war—another government plan that served only to distract us from the knowledge that we would have little control over our fates should such an event occur.

The Clinton vaccine plan—drawn up in haste and in reaction to renewed clashes with Iraq's Saddam Hussein over UN inspections of his purported bioweapons facilities—has a dubious future now that the dust of imminent threat has cleared. Although biotechnology could certainly take advantage of this short-sighted strategy, it has a responsibility—and a vested interest—to take the longer view and work toward the development of realistic plans that have at least a chance of being effective.

In this issue, Scott Layne and Tony Beugelsdijk present a proposal for "batch science," (see "Laboratory firepower for infectious disease research," p. 825) that could serve as the basis for more integrated solutions to dealing with multifaceted problems like bioterrorism. Batch science is an approach to solving problems that require large amounts of data analysis by linking laboratories over the Internet for problem solving both in real and nonreal time. Batch science machines can serve as

programmable laboratory technicians, performing the mechanical work of hundreds of human beings.

One example of batch science discussed by the authors is the typing of influenza A viruses by an automated reference library. To carry out various procedures, flu investigators would use a suite of process control tools to program manage and track procedures at every step. With respect to bioweapons, batch science facilities could be used to monitor, inspect, and test for infectious agents in a timely manner, to offer timely information in the event of an attack—How many agents were released? How do they differ? What is the best treatment for those affected?—and to assist in the aftermath, by identifying and categorizing lethal agents and helping to determine who made them. The authors believe that the creation of batch science facilities would not require a biological "Manhattan project," and that first-generation facilities could be up and running in 2–3 years.

A system such as the one they describe is an attempt at creating a coordinated plan for dealing with big problems—like biological warfare—that require big biological solutions. The medical community has long expressed its interest and involvement in finding solutions to problems posed by biological and biotechnology-driven threats. It's time for the biotechnology industry—which will make the vaccines and the antibiotics and the diagnostics that will protect and defend us from these threats—to do the same.

# Laboratory firepower for infectious disease research

Batch science machines could one day perform the work of hundreds of humans, enhancing public health surveillance, increasing experimental reproducibility, and liberating researchers from the bench so that they can focus on “big” conceptual problems.

Scott P. Layne and Tony J. Beugelsdijk

Infectious diseases pose “big” challenges for the biomedical community. New and lethal strains of influenza A virus are surfacing in Hong Kong<sup>1</sup>, pathogenic and drug-resistant food-borne infections are emerging in the United States and elsewhere<sup>2</sup>, and the number of HIV infections and AIDS cases worldwide is growing each year<sup>3</sup>. In addition, biowarfare and bioterrorism threats are looming—with the potential for destabilizing and disastrous consequences across the globe<sup>4</sup> (see “Beating the bioterrorist”)—and novel and uncharacterized infectious agents are emerging, the pathogenic properties of which remain only poorly understood<sup>5</sup>. Under such circumstances, infectious disease researchers must consider two related and important questions: First, how valuable is it to speed up the process of research that explores the very broad spectrum of diseases, antibodies, therapies, and vaccines? Second, if speeding up the research process is valuable, what are the feasible scientific and technical approaches to consider?

For the first time, a critical number of useful technologies and scientific disciplines can be brought together for leveling the playing field against the threat of infectious diseases. Powerful laboratory tests (methods), automation and robotics (hardware), object-oriented programming languages (software), relational databases (informatics), shipping services (virtual warehousing), and Internet providers (communications) are available to infectious disease researchers. The challenge is to select problems that demand focused attention and then to merge cross-disciplinary technologies for effective action.

Scott P. Layne is an associate professor in the Department of Epidemiology, UCLA School of Public Health, P.O. Box 951772, 73-320 CHS, Los Angeles, CA 90095-1772

([spl@lvik.ph.ucla.edu](mailto:spl@lvik.ph.ucla.edu)) and Tony J. Beugelsdijk is a staff member at Applied Robotics and Automation Group, Engineering Sciences and Applications Division, P.O. Box 1663, MS J580, Los Alamos National Laboratory, Los Alamos, NM 87545 ([beugelsdijk@lanl.gov](mailto:beugelsdijk@lanl.gov)).

## Digital versus physical

Unprecedented digital technologies are now at the fingertips of scientific and medical researchers in many geographic locations. At many advanced computer facilities, numerical performances are surpassing  $10^{12}$  floating point operations per second (teraflop speeds); database storage facilities are exceeding  $10^{15}$  bits per site (petabit capacities); and Internet communication rates are achieving  $10^9$  bits per second (gigahertz bandwidths). In fact, these values probably underestimate current state-of-the-art capabilities. Nevertheless, despite such fantastic digital firepower, many important scientific and medical research efforts currently are limited by sheer physical firepower. Besides persistent intellectual commitment and strokes of good luck, the biggest rate-limiting step is the ability to perform vast numbers of experimental tests in the laboratory.

Most laboratory experiments in science and medicine are performed by human hands, usually in combination with sprinklings of labor-saving devices. Although this semimanual approach leads to insights and breakthroughs, such laboratory work is

repetitive and exceedingly tedious. Even with an army of postdoctoral researchers and laboratory technicians, human hands are the limiting factor in generating the vast quantities of information that are required for solving complicated scientific problems.

Biologists have developed a variety of laboratory-based assays that are reproducible and readily adapted to large-scale efforts. Engineers have created innovative automation and robotic technologies that are capable of skyrocketing the number and variety of laboratory experiments. Computer scientists have refined Internet programming languages and database management systems that are providing the basic building blocks for improved environments in scientific collaborations. Physicists—driven by the need to share gargantuan amounts of data generated in high-energy particle physics at a few large accelerator facilities—were instrumental in the creation of the World Wide Web.

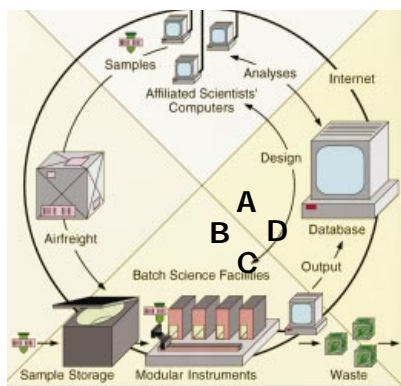
These developments are literally transforming the way in which scientific collaboration and information distribution is taking place. An integration of these capabilities

## Beating the bioterrorist

Mitigating the threats posed by biological weapons and terrorism will demand massive laboratory detection and characterization capabilities, which can be understood in terms of three overall phases:

- In phase 1, the ability to prevent an attack in the first place strongly depends on the ability to monitor, inspect, and test for certain infectious agents in a timely manner. Dedicated batch science facilities would contribute by being made available to responsible governments, United Nations weapons inspectors, and civilian scientists who wish to scrutinize suspicious occurrences and facilities<sup>4</sup>.
- In phase 2, following a biological attack, current infectious disease laboratories would be overwhelmed within the first few moments—quite simply because there would be too many samples to process and test from those afflicted. Simply put, so-called traditional laboratories would be incapable of answering even the most fundamental questions. How many different infectious agents were released? How do they differ? What are the best initial ways to treat those infected? Dedicated batch science facilities would contribute by offering timely information in acute situations<sup>10</sup>.
- In phase 3—the aftermath of an attack—public health and law enforcement officials would need accurate answers to yet another set of questions: What is the stability of each infectious agent? What are their geographic boundaries? What is the genetic fingerprint or origin of the agent(s)? Dedicated batch science facilities would contribute here by offering high-throughput laboratory capabilities for cleanup and investigative operations.

## FEATURE



**Figure 1.** Batch science via the Internet would enable researchers (at any geographic location) to submit sets of biologic samples and automated-instrument instructions as coordinated units. This approach is distinct from virtual science via the Internet, where experiments are controlled from afar in real time. (A) Scientists located anywhere in the world design experiments in cooperation with a particular batch science laboratory. (B) Shippers deliver packages containing bar-coded specimens or reagents. (C) Batch science laboratories house flexible, modular, and scaleable instruments. (D) Database facilities maintain permanent records and provide software for analyzing and managing data.

holds significant promise for accelerating infectious disease research, including basic science, clinical trials, and public health investigations throughout the world.

### Batch science

The Internet can be used by laboratory-based researchers in one of two ways: Either as a series of “real-time” or as a series of “non-real-time” operations. Real-time operations require specialized communication protocols that are redundant and fail-safe (i.e., they have large communication bandwidths). An illustration of real-time manipulations would be microscopy from a distance, whereby a microscopist operates the controls in one location and the sample is manipulated and viewed in another place. In fact, only a few specialized applications in science and medicine require real-time operation from afar. Moreover, the Internet’s wiring and communication protocols are not quite ready for handling real-time manipulations on an everyday basis. More capabilities and infrastructure would be required beforehand.

Non-real-time operations, however, do not require specialized communication protocols that are redundant and fail-safe (i.e., they have small communication bandwidths). Most scientific and medical efforts can be supported by non-real-time operations that are scripted with flexible software tools—an approach that is referred to here as batch science via the Internet. In this scenario, batch science machines would serve as programmable laboratory technicians, perform the mechanical work of hundreds of humans, and help infectious disease researchers tackle “big” problems.

An illustration of batch science would be the typing of influenza A viruses by an automated reference laboratory (Fig. 1). In this example, teams of investigators would collect and bar code numerous infectious samples around one province (e.g., at the local health authorities in Hong Kong) and then use flexible software tools to design and script their flu-typing assay soon after. Next, authorities would airfreight the frozen flu samples to their

collaborating facility (e.g., at the Centers for Disease Control and Prevention, Atlanta, GA), and within just a few days the desired assays would be set up and carried out by high-throughput automation in accordance with “assay scripts” that arrived via the Internet.

Upon completion of the work, data from the numerous typing assays would be deposited electronically into the international influenza database (at the World Health Organization

[WHO] in Geneva) and collaborating teams of investigators would incorporate this information into their ongoing worldwide surveillance and planning efforts. When compared with flu-typing assays performed by human technicians, the quality assurance team (at the National Institutes of Health, Bethesda, MD) would further observe that batch science methodologies are significantly faster, offer larger sample sizes, and exhibit greater reliability, all of which would translate into far better preventive strategies against emerging and lethal strains of influenza A<sup>6,7</sup>.

To carry out the various procedures illustrated above, flu investigators would use a suite of process control tools (PCTs; Table 1) to program, manage, and track scientific procedures at every step. For instance, PCTs would be used to assign bar codes to samples, script individualized assay protocols, analyze the raw data, and create relational links with associated information. Also, at every step along the way, private or sensitive information would be protected by any number of accepted encryption and authentication methods. This too would be handled by PCTs (see “Getting connected”).

## Getting connected

Process control tools (PCTs) would be needed to connect investigators from any geographic location to batch science machines and their associated database facilities. PCTs would be designed to permit maximum flexibility and control over experiments by remote investigators—just as if they were employing laboratory technicians to carry out their assay protocols. To simplify their use, PCTs would also be designed to interface easily with World Wide Web browsers (e.g., Microsoft Explorer, Netscape Navigator, and Sun Hotjava) and data analysis software packages that are commercially available. With object-oriented programming languages for the Internet, these PCTs would be designed to accommodate improvements over time and support practically any type of application. To enable just about any type of research activity, one can envision a collection of seven basic PCTs (See Table).

**Table 1.** The basic types of PCT.

PCT type	Operation
Access	Enable appropriate authorization and gateway functions.
Operation	Describe how to use batch science machines, offer a selection of standardized laboratory tests, and enable investigators to compose their own assay scripts.
Documentation	Performs a variety of annotating functions, enabling investigators to deposit relevant background information on specimens, reagents, and related matters. To afford maximum flexibility, the formats could include written, audio-, and video-based documents.
Submission	Tells investigators how to package their specimens and associated reagents, and generates identifying bar codes to be affixed to all containers before shipping. To support the high-throughput environment, every incoming item would be packaged in standardized containers (tubes, bottles, etc.) so that they can be fed directly into SLM-based machines.
Analysis	Provides parsing tools for manipulating raw data formats, computational tools for analyzing data, and relational tools for linking data to other types of information. It would also provide convenient links to corresponding quality-control data from the batch science facilities.
Privileges	Allows submitting investigators to designate who has permission to view or use their data. Feasible options for these information management arrangements include: First, access by submitting investigator only; second, time-embargoed data followed by wider access; third, access by certain designated collaborators; and fourth, unrestricted access by all.
Commerce	Deals with the business aspects of batch science facilities such as audit trails, billing services, inventory management, and cost modeling.

### Modular, flexible, and scaleable

Specialized and automated laboratory modules are available for conducting practically every necessary task in infectious disease research—such as bar coding, liquid handling, centrifuging, incubating, sequencing, immunostaining, and image capturing. However, most major companies have developed unique software interfaces for their high-throughput laboratory automation and robotic products. This fragmentary environment is workable, as long as customers purchase products from a single company, which happens only rarely in most laboratory settings. Thus, a real frustration for researchers is that it takes an inordinate amount of effort to incorporate components from different companies into one effective instrument.

The American Society for Testing and Materials is now in the process of adopting the Laboratory Equipment Control Interface Specification (LECIS), which defines the handshake between any number of standard laboratory modules (SLMs) and one task sequence controller (TSC; Fig. 2). LECIS would enable laboratory automation and robotic companies to take advantage of a single standard for interconnecting all types of commercial hardware and lead to automated laboratory modules that exhibit many of the compatibility conveniences found in personal computer-based components<sup>8</sup>. From a design standpoint, SLM-based machines house groups of self-contained modules, with each SLM dedicated to a particular set of tasks. Suppose that the first task in the flu-typing assays used in the example above is to thaw the sample tubes from Hong Kong, open the tubes, and then add fresh culture media. In this case, a liquid handling module would perform the set of tasks outlined below under the overall command of the TSC:

```
SLM ready→Input capped/frozen
tube→Inspect→Warm to
37°C→Uncap→Add liquid→Re-
cap→Vortex→Verify
weight→Output capped/recon-
stituted tube→SLM ready
```

There is no upper limit to the number of modules per machine, or to their physical size. In practical terms, of course, the actual number of SLMs depends on the particular research application and throughput requirements. Analogous to computer architecture, duplicate modules can be used to add parallel-processing capabilities to machines, to build batch science machines based on conventional macroscale technologies (pipettes, test microarrays, etc.), newer nanoscale technologies (microchannels, microchambers, microarrays, etc.), or modular combinations of both. A strategic assortment of merely 10–20 different SLMs could satisfy most infectious disease research needs, facilitating the rapid configuration and reconfiguration

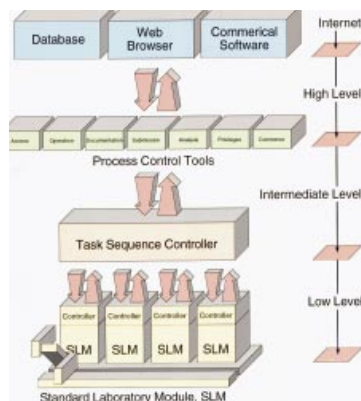


Figure 2. Batch science machines would operate on a three-tier hierarchy. At the lowest level, modules possess SLM controllers that drive components like actuators, detectors, and servomotors and coordinate their internal electromechanical activities. Such events are contained wholly within each module (i.e., modules are blind to the existence of one another), which gives rise to independently working SLMs. At the intermediate level, task sequence controllers use tools from operations research to govern intricate flows of reagents and samples through the entire machine. Various standardized commands and feedback signals work to carry out complete laboratory procedures or programmable assay scripts. At the highest level, process control tools would enable scientific and medical researchers to carry out a spectrum of important activities via the Internet.

of high-throughput research in a highly flexible operating environment. It now becomes an issue primarily of the bigger strategies and directions that researchers wish to engage.

### Scientific needs

There are many compelling applications for batch science via the Internet. To determine the most promising initial applications, one could choose research problems that impact large numbers of people around the world;

require large quantities of data for achieving solutions; afford access to appropriate numbers of samples for testing and analysis; leverage reproducible laboratory procedures; or offer sufficient quantities of supporting reagents. Some potential applications are listed in Table 2.

For each problem shown in Table 2, dedicated clusters of batch science machines (in one or more locations) could serve the needs of investigators throughout the world. For example, the typing of influenza viruses would be accomplished with three batch science machines that work in series. The first machine would culture viral isolates from various bar-coded samples collected in the field. The second machine would harvest the viral cultures and prepare various arrays of specimens for molecular sequencing. The third machine would perform short-length or full-length PCR analysis. (Engineering note—for the sake of simplicity, maintenance, and flexibility, it is better to build several smaller machines instead of one larger unit.) From the investigator's standpoint, all three flu-typing machines would work together as a single integrated unit and take advantage of economies of scale (Table 4).

In various fields of research, investigators often rely heavily on small assortments of different laboratory methods or tests. For example, most virologic work (on pathogens like influenza A, hepatitis C, HIV, etc.) currently is carried out with an experimental "toolbox" that consists of infectivity assays, cytotoxic cellular assays, and molecular sequencing assays. These different laboratory tests are then customized and optimized to meet the particular needs at hand. For each specialized application, designers would consider the most strategic way to build a small cluster of SLM-based machines that work together (i.e., in series or parallel) to serve investigators' needs. The bottom line being that small clusters of high-throughput machines would be flexible enough to cover most laboratory experiments in particular fields.

Table 2. Biological applications for batch science.

Category	Focus
Infectious	Detecting and rapidly characterizing biological warfare agents. Intensifying global warning programs against lethal strains of influenza A. Characterizing rapidly emerging infections, such as hepatitis C. Surveying for drug-resistant infections, such as tuberculosis and malaria. Testing new HIV/AIDS therapies and vaccines.
Cardiovascular	Discovering therapies for circulatory diseases, such as atherosclerosis. Discovering therapies for cerebrovascular diseases, such as strokes.
Cancer	Understanding genes expressed in cancer versus normal cells. Determining how environmental factors activate cancer genes, such as <i>p53</i> . Evaluating new anticancer drugs and combinations of therapies.
Agriculture/ environment	Genetically engineering new varieties of high-yield crops. Protecting food supplies against pathogens, such as <i>Escherichia coli</i> O157:H7. Monitoring environmental pollutants, such as <i>Pfiesteria piscicida</i> toxins.

The list emphasizes infectious diseases research mainly because infectious agents exact a great toll on human life and are capable of causing even greater suffering than currently observed<sup>9,10</sup>.

## FEATURE

**Research infrastructure**

Historically, researchers working at global pharmaceutical companies have had ample access to integrated high-throughput automated laboratory instruments, whereas

researchers working at most major universities have enjoyed far less access to such experimental firepower. This lopsided situation, however, is ripe for change. In order to maintain a pipeline for creating new therapeutics,

global pharmaceutical companies are forming increasing numbers of alliances with external academic research groups<sup>16</sup>.

Consistent with the trend in research "outsourcing" batch science facilities would thus serve as national and international centers for carrying out investigator-initiated research, would be sponsored by consortia of government agencies and research-driven companies, and would be located at universities or national laboratories. During the start-up period, the center would receive major financial support from sponsoring organizations, permitting the creation and validation of its initial resources. After the center became fully operational, the level of start-up support would be scaled down and, over several years, feasibly evolve into a nonprofit entity that charged investigators reasonable fees for "mass customized" testing services.

By offering the same common resources to participating investigators, batch science facilities would help to eliminate many of the problems associated with comparability. The high-throughput environment would permit enormous numbers of experiments to be run in parallel, and against a series of standards to certify reliability. Also, by offering a series of standardized laboratory protocols (in the form of digitized assay scripts), the laboratory facilities would help to eliminate variable outcomes caused by minor changes in experimental protocols.

All too often, researchers shy away from the problems that deal with the "bigger picture" simply because they lack the necessary tools to move ahead. In HIV/AIDS research, for instance, many important problems seem beyond reach because of the enormous "experimental spaces" to be explored and characterized. The HIV RNA genome comprises approximately 10<sup>4</sup> bases, corresponding to about 3000 amino acids. If only a small fraction (about 1%) of amino acids were responsible for conferring unique viral properties, and only one specific amino acid (from the total selection of 20) conferred this property, there could be upwards of 2<sup>30</sup> ≈ 10<sup>9</sup> unique viral genotypes. This is truly an enormous number to attempt to sample and characterize.

At present, we have only sketchy data relating genotypic variations to phenotypic expression, making it difficult to relate the significance of such variations (if any) to therapies and vaccine development. What little knowledge we do have is hopelessly scattered among various investigators' notebooks. Batch science facilities, with the capacity to perform the work of hundreds of laboratory technicians, would enable investigators to conceptualize and attack important problems from entirely new directions. For many such problems with large phase spaces, high-throughput laboratory facilities with digitalized record keeping are perhaps the only feasible means to move ahead.

**Rewarding batch science**

With high-throughput technologies changing the means by which research is conducted, it will be important to maintain the traditional reward system for investigators. At the heart of this system is the freedom to decide how to share data and new information, which can lead to scientific publications and credit for discoveries involving intellectual property<sup>11</sup>. Several categories of data ownership and access can be envisioned according to the source of financial support (Table 3). For the closed category, data would belong solely to the commercial organization that submitted samples/assay scripts and paid for the research or testing services. Upon completion, the batch science facility would encrypt and forward all the raw data to that purchasing organization. (The facility would also temporarily maintain a secure copy of the digital records to assure redundancy and integrity in accordance with contractual agreements.) In situations in which either a single principal investigator or a collaboration of researchers is receiving government grant support, data would be retained for a reasonable period of time after grant expiration (2–3 years). After this time-embargo had expired, relational links would be attached to the digital records and the information would become available to others. In the case of single investigators, good digital practices would be tied to ongoing grant support, requiring that database records be maintained in an orderly manner. Digital records would also have to be organized carefully in consortia, most likely under the supervision of the group's database manager. In the last category, an open arrangement, data and associated links would be accessible to the public at all times. The digital records would come from voluntary submissions and time-embargoed data, which would be released automatically. The main issue would be maintaining backup copies to assure integrity, as well as deciding how to collate the data and build relational links. The National Center for Biotechnology Information (Bethesda, MD), the European Molecular Biology Laboratory (Heidelberg, Germany), and the DNA Databank of Japan (Mishima) are examples of public database facilities in the United States, Europe, and Asia, respectively<sup>12–14</sup>. These organizations have already established guidelines regarding time-embargoed ownership of data, which could provide reasonable starting points for standard agreements pertaining to batch science<sup>15</sup>.

**Table 3. Data ownership and privileges**

Category	Funding	Data access
Closed	Commercial organization	Organization maintains sole ownership of the data.
Principal investigator (PI)	Government agency supports single PI	Investigator has time-embargoed ownership of the data
Consortia	Government agency supports many PI's	Collaboration has time-embargoed ownership of the data
Open	Many sources of support	Anyone can use the data any time

**Table 4. An economic comparison of a manual laboratory to a batch science facility with the same throughput capacity.**

Manual facility	Cost (millions)	Batch science facility	Cost (millions)
<i>One-time capital costs</i>			
Equipment	\$10	Automation equipment	\$5
50,000 sq. ft floorspace	\$10	5,000 sq. ft floorspace	\$2
<b>Total</b>	<b>\$20</b>	<b>Total</b>	<b>\$7</b>
<i>Annual operating costs</i>			
100 technicians salaries	\$5	10 technicians' salaries	\$0.5
40,000 assays	\$10	40,000 assays	\$2
Maintenance	\$2	Maintenance and upgrades	\$2
<b>Annual total</b>	<b>\$17</b>	<b>Annual total</b>	<b>\$4.5</b>
	× 5 = \$85		× 5 = \$22.5
<b>Cumulative 5-year total*</b>	<b>\$105</b>	<b>Cumulative 5-year total*</b>	<b>\$29.5</b>

Batch science facilities would take advantage of economies of scale, thereby reducing the unit cost of laboratory experiments by over threefold (unit cost per assay in a manual facility ~\$25 compared with ~\$150 in a batch science facility). The overall savings would make it possible for investigators to carry out larger experimental undertakings for the same dollar expenditures. \*Five year total is sum of first year's capital cost and five years of operating costs.

Health care is progressing in the direction of procedures that are precisely tailored to individuals or smaller groups of patients. Some of the most promising initial applications include infectious disease treatments, customized cancer chemotherapy, and transplant-based genetic therapies<sup>17</sup>. For instance, AIDS patients who take antiviral drugs must rely on just the right combination (i.e., reverse transcriptase inhibitors and protease inhibitors) to halt viral replication. With current clinical and laboratory practices, however, there is no single test to establish the optimal regimen.

Thus, to maintain effective therapies, physicians have come to rely on a series of tests (i.e., viral loads, molecular sequencing assays, and drug susceptibility assays) that are highly repetitive, time-consuming, and labor intensive. Batch science facilities would assist by both providing the necessary high-throughput testing services and creating an organized base of information for clinical decision making. In addition, pharmaceutical companies could use batch science facilities to track the molecular evolution of HIV in patients around the world. A wider base of information could then be used to guide the development of new generations of antiviral drugs.

In the developing world, scientists are often forced to work with primitive laborato-

ry resources. Ironically, for infectious disease researchers, this is precisely where meaningful studies should be conducted. With limited infrastructure in the field—an Internet connection, plasticware, bar codes, refrigeration, and air transportation—one batch science facility could serve large geographical regions, facilitating research.

The creation of batch science facilities will not require a “Manhattan Project” (or major redirections of resources) as some might suspect. On the contrary, it will involve a focused effort by a relatively small team of motivated engineers and scientists. With appropriate support, the first-generation facilities could be running within two to three years, and next-generation refinements would follow soon after. If we assume that the facilities perform the work of say 100 laboratory technicians at a fraction of the cost, the development costs would be recouped within several years.

The time has come for government funding agencies, the WHO, health-orientated foundations, and global pharmaceutical companies to consider their respective roles in building “intermediate-scale” research infrastructures<sup>19</sup>. Scientific and medical researchers, particularly in the area of infectious diseases, will certainly have no problems finding clever (and unforeseen) ways to use them for a spectrum of human diseases.

1. Belshe, R.B. 1998. Influenza as a zoonosis: How likely is a pandemic? *Lancet* **351**:460–461.
2. Tauxe, R.V. 1997. Emerging foodborne diseases: An evolving public health challenge. *Emerg. Inf. Dis.* **3**:425–433.
3. Heilman, C.A. and Baltimore, D. 1998. HIV vaccines—where are we going? *Nat. Med. Vaccine Suppl.* **4**:532–534.
4. Lederberg, J. 1997. Infectious diseases and biological weapons. *J. Am. Med. Assoc.* **278**:435–436.
5. Gore, A. 1996. Emerging infections threaten national and global security. *Am. Soc. Microbiol. News* **62**:448–451.
6. Yuen, K.Y. et al. 1988. Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *Lancet* **351**:467–471.
7. Class, E. et al. 1998. Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet* **351**:472–477.
8. American Society for Testing Material, Subcommittee E-49.52. 1998. Laboratory equipment control interface specification (LEICIS): <http://www.thermal.esa.lanl.gov/astm>.
9. Beugelsdijk, T.J. 1989. Trends for the laboratory of tomorrow. *J. Lab. Robot. Autom.* **1**:11–15.
10. World Health Organization. 1993. Global health situation III, mortality. *Weekly Epidemiol. Rec.* **68**:33–36.
11. National Research Council. 1997. Intellectual property rights and research tools in molecular biology. National Academy Press, Washington, DC.
12. <http://www.ncbi.nlm.nih.gov>
13. <http://www.embl-heidelberg.de>
14. <http://www.ddbj.nig.ac.jp>
15. Layne, S.P. et al. 1998. The need for national HIV databases. *Nature* **333**:511–512.
16. Herrling, P.L. 1988. Maximizing pharmaceutical research by collaboration. *Nature Suppl.* **392**:32–35.
17. Anderson, W.F. 1998. Human gene therapy. *Nature Suppl.* **392**:25–30.
18. National Research Council. 1998. Improving civilian medical response to chemical or biological terrorist incidents. National Academy Press, Washington, DC.
19. National Research Council. 1993. National laboratories: Applying information technology to scientific research. National Academy Press, Washington, DC.