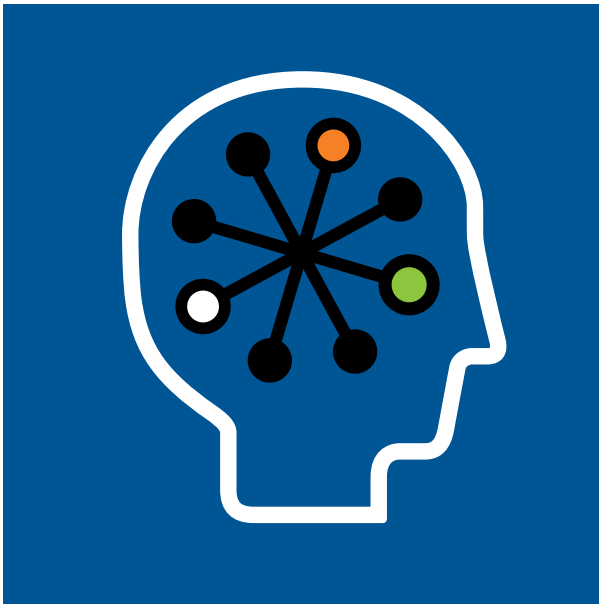


Extending the IT foundation for personalized medicine

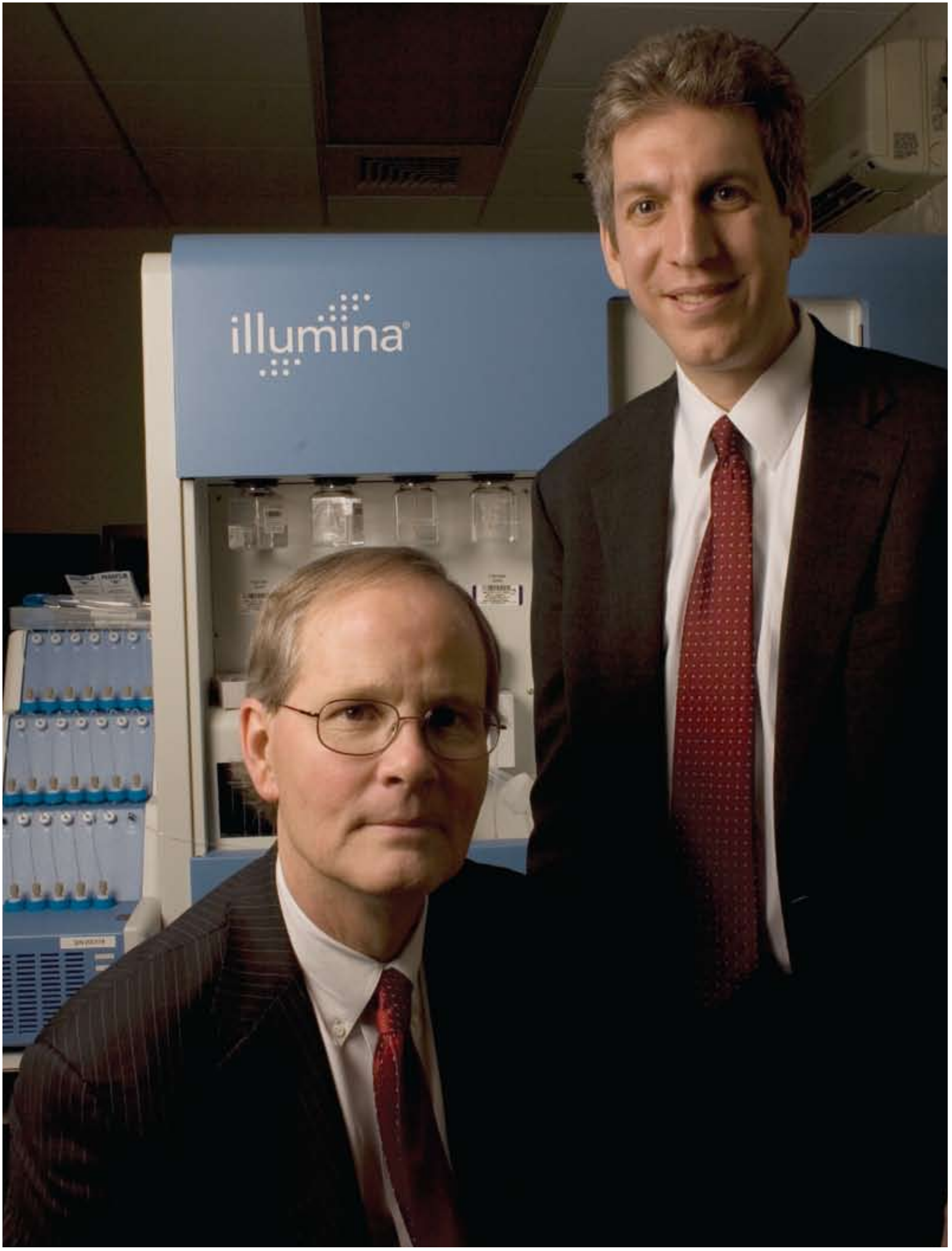
Harvard Medical School — Partners HealthCare Center for Genetics and Genomics wanted its storage to keep pace with data volumes from new sequencing technologies.

In terms of health care IT, what does the next five years look like? It's an open-ended question, but the IT team at Harvard Medical School – Partners HealthCare Center for Genetics and Genomics (HPCGG) has a quick and targeted response. To support its vision of personalized medicine, where clinicians are able to make treatment and medication decisions based on a patient's genetic profile, IT has and will continue to play a huge role.

In the past five years, HPCGG focused its efforts on building a viable IT foundation with the right hardware and innovative software. "Making personalized medicine really work requires a vast amount of IT infrastructure," says Sandy Aronson (pictured on right, next page), Executive Director of Information Technology, HPCGG. "We've been building that foundation incrementally so we are in a strategic position to address the challenges stemming from next generation sequencing technologies."

In partnership with HP, HPCGG deployed a large multi-cluster compute facility to support researchers throughout Partners HealthCare, which includes Massachusetts General Hospital and Brigham and Women's Hospital. In addition to processing capabilities, HPCGG implemented a shared storage approach for raw instrument data, results files and database objects for increased collaboration across institutions.

Leveraging that shared infrastructure, HPCGG and HP collaborated in the creation of the Gateway for Integrated Genomics-Proteomics Applications and Data (GIGPAD). The software environment integrates disparate laboratory information management systems (LIMS), provides a portal to submit and access biological samples, and enables efficient management of analytical workflows and data repositories.


illumina

"We work closely with HP to constantly develop, enhance and deploy the IT solutions that enable us to keep pace with the field."
John Glaser, CIO
Partners HealthCare

By keeping the biomedical and life science data in one place, users can work on it almost immediately instead of waiting a day for the migration.

"Everything we do is to benefit the patient," says Trung Do, Executive Director Business Development, Partners HealthCare. "While the deeper integration of genetics and genomics into clinical practice is a huge advance for patient care and diagnostics, we also recognized the pressure it was exerting on our systems. There was a new set of requirements in how we practice medicine and run our business, and we had to ensure that IT remained an enabler of both research and patient care."

IT as an enabler

For organizations like HPCGG, there is an ongoing concern that IT must keep pace with the biology, especially in terms of storage capabilities. As the costs of sequencing technologies continue to fall and the volumes of data from those technologies continue to rise, that concern becomes a reality that IT must grapple with daily. The focus is on ensuring that IT is not a constraining factor in realizing the benefits of next generation sequencing technologies.

"The rate of change in the genomics field is accelerating with new instruments initially introduced for research purposes transitioning into clinical use," says John Glaser (pictured on left, previous page), CIO, Partners HealthCare. "We are seeing testing volumes on these new instruments ramp very rapidly. We work closely with HP to constantly develop, enhance and deploy the IT solutions that enable us to keep pace with the field."

With the new technologies, it can take approximately one month to sequence a human genome in a research context, which is about 10x faster than what was available previously. The sequencing process generates very large data volumes, and the complexity associated with image analysis, quality scoring, base-calling and information analysis requires a high performance computational environment.

Using the previous generation of sequencing technologies, organizations could reasonably predict data growth; estimates hovered around a doubling of data every 18 months. But the new technologies have changed all of that.

"Four years ago, we were able to have a fairly accurate picture of our storage needs," says Brent G. Richter, Director of Enterprise Research Infrastructure and Services, Partners HealthCare. "At the end of 2007, we expected to have approximately 10 terabytes of data across all functions in the organization. But then we obtained two next-generation sequencing machines in that year, and far exceeded those estimates. By the beginning of 2010, we now expect a petabyte will be required to be online."

HPCGG plans to acquire one more new sequencing machine in the next three months, with a fourth in the six to nine month timeframe. Adding to its storage requirements, HPCGG is also responsible for supporting other research platforms and technologies across the Partners HealthCare teaching hospitals. According to Glaser, "Rapidly decreasing sequencing costs have the potential to increase progress in the field of personalized medicine. However, significant IT infrastructure will be required to realize this potential."

Richter highlights that the falling costs of sequencing will lead to increased demand—more labs, researchers and organizations will see genomics as an attractive and viable option for research and diagnostics. And that will increase the pressure on IT resources to act as an enabler rather than an obstacle.

Cost-effective storage

Many organizations choose to manage the storage challenge by discarding the principal data (images) once researchers get to the second level of analysis. But there are implications to throwing away this data when biological samples are hard to obtain and replace. The key is to find a way to store and share that data cost-effectively from the initial analysis of generating base-call sequence data and the analysis by investigators and diagnostics laboratories through to post analysis visualization and dissemination of results to researchers, electronic medical records or clinicians.

To address its storage-related issues, HPCGG chose HP Extreme Data Storage, which provides multi-protocol file service backed with a half petabyte of usable storage. In addition to the cost-effective storage capacity, HPCGG will be able to integrate the new storage system with its existing cluster environment so users can work on the data without moving it through the network.

"Moving large amounts of data around is a significant challenge for us," says Richter. "When we have to move data from Modular Smart Array file storage to the High Performance Computing storage environment, it generates a bottleneck in the delivery of results to end users. By keeping it all within the same storage pool we can eliminate the migration, save on network bandwidth and reduce overall processing time."

And that is a significant benefit for researchers. For example, an entire genome generates approximately 30 terabytes of image data. Investigators or molecular labs end up working with one or two terabytes of that data for follow-on analysis. For HPCGG, the general benchmark for moving data is one day for one terabyte. By keeping the biomedical and life science data in one place, users can work on it almost immediately instead of waiting a day for the migration.

"Our ultimate goal is to enable clinicians to incorporate the latest genetic information into patient care so that they can maximize the 14.7 minutes they typically have with each patient," says Aronson. "Working with HP, we are building and extending our IT foundation to realize our vision of personalized medicine—ensuring genetic and genomic discoveries flow quickly through IT and governance so they can benefit patients at the point of care."

This article first appeared in the Spring 2009 "Driving breakthrough quality in health and life sciences" supplement to HP's "Transforming Your Enterprise" Magazine. The current edition of this publication is available at: www.hp.com/go/transformHLS

For more information, see HP StorageWorks at:
www.hp.com/go/transformHLS

4AA2-0000ENA March 2009

invent