White Paper
Intel® Xeon® Processor E5 Family
High-Performance
Computing Healthcare

Delivering Personalized Medicine Today

How TGen, Dell, and Intel are Working to Make Personalized Medicine Mainstream

Executive Summary

Personalized medicine—marked at times by both excessive hype and dour skepticism—is a reality today, albeit one with a still-modest reach. One of the best examples is an FDA-approved, first-of-its kind pediatric clinical trial[1] in which children with a deadly, fast-growing cancer (neuroblastoma) have their tumors biopsied and characterized by modern molecular medicine tools (microarrays and next-gen sequencing). The resulting RNA profiles (gene expression) were interpreted by a predictive analysis platform that matched promising drugs to each patient's individual tumor.

Speed is crucial because trial subjects were either refractory or relapsed with poor prospects. These are children, mostly under age five. The trial goal was to complete biopsy-to-treatment determinations in 21 or fewer days. A multidisciplinary Molecular Tumor Board,[2] informed by the computational predictions and its own expertise, made final therapy decisions. So successful was this first proof-of-concept trial with just 14 patients that a follow-on trial[3] encompassing all pediatric cancers in a much larger (100-plus) patient cohort, has begun, with patient recruitment now underway. Consider the first trial's encouraging results:

- "Our primary endpoint was to see if 75 percent of the patients could be treated using this [process] and we actually surpassed this. All 14 were able to be treated using the molecular guided therapy and we found no serious or adverse side effects," reported Dr. Giselle Sholler, Chair, Neuroblastoma and Medulloblastoma Translational Research Consortium, Haworth Family Endowed Director of Innovative Therapeutics Clinic, Head Pediatric Oncology Translational Research Program, Pediatric Oncology, Helen DeVos Children's Hospital, and Associate Professor Michigan State University and lead sponsor of the clinical trials.[4] Dr. Sholler's remarks were given at a 2013 NMRTC symposium.
- "In terms of responses, one patient had a partial response [and] eight of the 14 had prolonged, stable disease, so over 60 percent did benefit clinically from this study," said Dr. Sholler. What's more, after analyzing one tumor's RNA profile, the predictive analysis platform suggested a drug not previously used against neuroblastoma. It proved effective in that particular patient, but not in others, reinforcing the power of personalized medicine and of the trial platform and approach.

This is personalized medicine, making a difference today with the promise of wider impact in the future, and deeply enabled by high-performance computing (HPC). Without HPC infrastructure, domain expertise, and financial support from Dell, as well as application expertise from Intel, this clinical trial (Molecular Guided Therapy for Refractory or Recurrent Neuroblastoma[5]) would have missed one of its primary objectives—demonstrating that next-gen sequencing, a key enabler for personalized medicine, is a viable tool in a time-constrained clinical setting.

Practically speaking, most of modern molecular medicine requires HPC. The datasets are too large to be handled otherwise. Files from sequencing a single tumor, for example, can exceed 700 gigabytes (GB). Managing and analyzing this data is often described as a perfect storm on the back end for IT and informatics. Dr. Sholler noted early in the trial, "Sequencing itself takes about two weeks, but the [data] analysis takes time too. Right now, two months is the quickest we can get good data to make clinical decisions."[6]

Dell Joins TGen and the Trial Effort

In November 2011, as part of its corporate responsibility commitment, Dell announced a relationship with TGen to help create a scalable HPC infrastructure that could process next-gen sequence data and deliver quantitative RNA profiles on par with microarrays.[7] (In the first trial, microarray data produced in a **CLIA** setting provided the decisive information for the Tumor Board. TGen worked in parallel on next-gen sequencing in a research setting. TGen has since obtained CLIA certification, enabling use of richer, next-gen data for clinical decisions in the second trial.)

"Our target was to turn around results, from biopsy to bioinformatics report, in roughly 12 days. That is approximately the time needed to perform a similar analysis of gene expression using microarrays," says Matt Huentelman, PhD, associate professor, TGen, in charge of the sequencing and the data analysis pipeline. "Our overall goal was to enable the next-gen sequencing approach to become a feasible competitor to array technology. Although next-gen sequencing generates richer data with an expanded dynamic range, it just wasn't fast enough to be used in the management of rapidly-progressing disease."

Working with Dell, Intel, and other technology leaders, TGen was able to streamline sample preparation and sequencing from 12 days to five, slash data analysis from several days to six hours, and demonstrate the feasibility of delivering results to the Tumor Board in roughly 10 days.[8] This is a remarkable achievement, due for the most part to careful planning and deployment of a well-designed HPC infrastructure including Lustre* storage, Infiniband* networking, and 28 TFLOPs of Intel® processors.

"We've gotten far more than iron boxes that run fast processors from Dell and Intel," says Dr. Jeffery Trent, president and scientific director, TGen. "We've had the help of their thought leaders on issues such as how to manage these massive datasets and unparalleled access to novel technology including Intel® Xeon® processor-based servers. I think this is one of the most successful examples of the intersection between industry and academia providing a framework that others can employ."

Review of the TGen-Dell-Intel collaboration spotlights computational challenges commonly confronted in the life sciences and the need for a multidisciplinary approach to solve them. At TGen, the focus was demonstrating the suitability of next-gen sequencing for personalized medicine. Technology choices around computer hardware (Dell PowerEdge* servers), storage and file systems (Lustre*), and cluster management (Bright Cluster Manager*), networking (InfiniBand*), and bioinformatics (Bowtie,* TopHat,* and Cufflinks*) were all part of the project.

"We knew we needed a lot of processors and a fast storage system to push data to all of these processor cores as quickly as possible," says James Lowey, vice president of technology, TGen. "Also, our data center was pretty much out of space and power just running existing programs. We had to look for an outside collaborator to house the system."

Lessons from the TGen-Dell-Intel Collaboration

Building the computational infrastructure from scratch enabled TGen to maximize Dell's full range of capabilities. "We are not just about putting different technology roadmaps in front of people," says Glen Otero, PhD, life sciences HPC solution architect, research computing solutions, Dell. "We have a set of best practices in many different areas, particularly in genomics, and can help the customer choose the right technologies, along with testing and building the appropriate solution at our facility before we put it in front of the customer. This is particularly valuable in the life sciences and healthcare, where HPC expertise is often limited."

Major planning areas included:

- Computational power (and cluster management)
- Fast storage and networking
- New data center
- Accelerated bioinformatics
- Sequencing efficiency
- Trial (and future regulatory) compliance

Computer hardware selection required extensive benchmarking at Dell's test lab. Several Dell PowerEdge servers (M620, M420, and R720) were evaluated. The Dell PowerEdge M420 Blade Server emerged as the top choice. It provided the most gigaflops per rack and was either first or second in energy efficiency tests.

"We could fit 512 cores (32 nodes) into a 10U chassis. Intel had just released its Intel Xeon processor E5 family at about the same time. Having that technology and packaging meant we had a huge amount of computational power in a relatively small space, while not consuming enormous amounts of electricity," says Lowey. Energy was a critical issue because the new hosting data center offered generous terms, but linked to energy consumption.

To a fair degree, TGen requirements reflect those characteristic of genomics research and healthcare, says Dr. Otero. "Customers in other spaces—like fluid dynamics, for example—would insist on platforms with higher-speed processors, the higher wattage processors, but that's not what this space needs. The most appropriate genomics solutions blend power and efficiency."

Predictably, data I/O and networking were problematic bottlenecks, primarily because the new data center would be offsite. The Lustre file system was selected. Originally designed for supercomputing, Lustre is a technology with which TGen was already familiar. Most importantly, Lustre can push a lot of data, very quickly, to a larger number of nodes. InfiniBand technology was pressed into service based on its ability to provide the bandwidth necessary for doing multiple samples at the same time.

"One aspect that's a little different in this installation," says Lowey, "is we're using a 40 Gb fabric in the backend exclusively for storage. We wanted to be able to push the Lustre file system through the InfiniBand fabric. This enabled us to push about 9 Gigabytes per second (GBps) total throughput, which relieved some of the I/O bottlenecks.

"Another interesting piece of technology deployed was a CIFS* gateway, which the next-gen sequencers need to write data directly to the Lustre file system. Previously, sequencers had to write to NAS* systems, and then users had to copy that data to the Lustre system.

Obviously, that takes time, because files coming off the sequencers can be terabytes in size. Being able to write directly onto the Lustre file system is a huge time saver." Dell can provide Lustre HPC storage systems (HSS) into the petabyte range.

Although the new data center is only six miles away, the total fiber optic connection is approximately 75 miles and uses a 10 Gbps Ethernet link. TGen worked with another technology leader, Obsidian Technologies, to stretch the InfiniBand fabric through that 10 Gbps link between the two buildings. Much of the data traffic going through the link is encrypted but moves without performance degradation, says Lowey. Typically between 200 Mbps to nearly 800 Mbps move through the link—all AES 256 encrypted.

Such rigorous encryption is not required today—genomic data without patient identifiers is not considered protected health information (PHI)—but likely will be required in the future. "I believe when you build systems and make huge investments of this caliber, you want to build it toward the future," says Lowey.

Determining the system's energy requirements was also a challenge. "We were trying to build the system in a short time, and a system like this has many parts. When you start to work on the power calculation, you generally have all these unwieldy spreadsheets. Sitting down with Dell engineers and Dell tools was critical to success."

Bright Cluster Manager from Bright Computing was selected to manage the cluster. "When we compared it to Platform ECM* and Rocks* for cluster management, we found that Bright Cluster Manager was easier to use, much more scalable, and had more future-proof features. For example, it's able to burst into different types of clouds, whether it's OpenStack* or Amazon.* They were also willing to work with us to integrate some of our appliances to be recognized in their toolkit," says Dr. Otero.

"We did a lot of testing in our labs and have received positive feedback from customers. Our practice now is to lead with Bright Cluster Manager as the solution for cluster management," says Dr. Otero. "TGen was already using TORQUE* as a job scheduler for sequencing and wanted to stick with that, so it worked well." Once built, the new system was made available to all TGen researchers, but priority queuing ensured RNA-profiling analysis jobs for the trial were always bumped to the head of the queue.

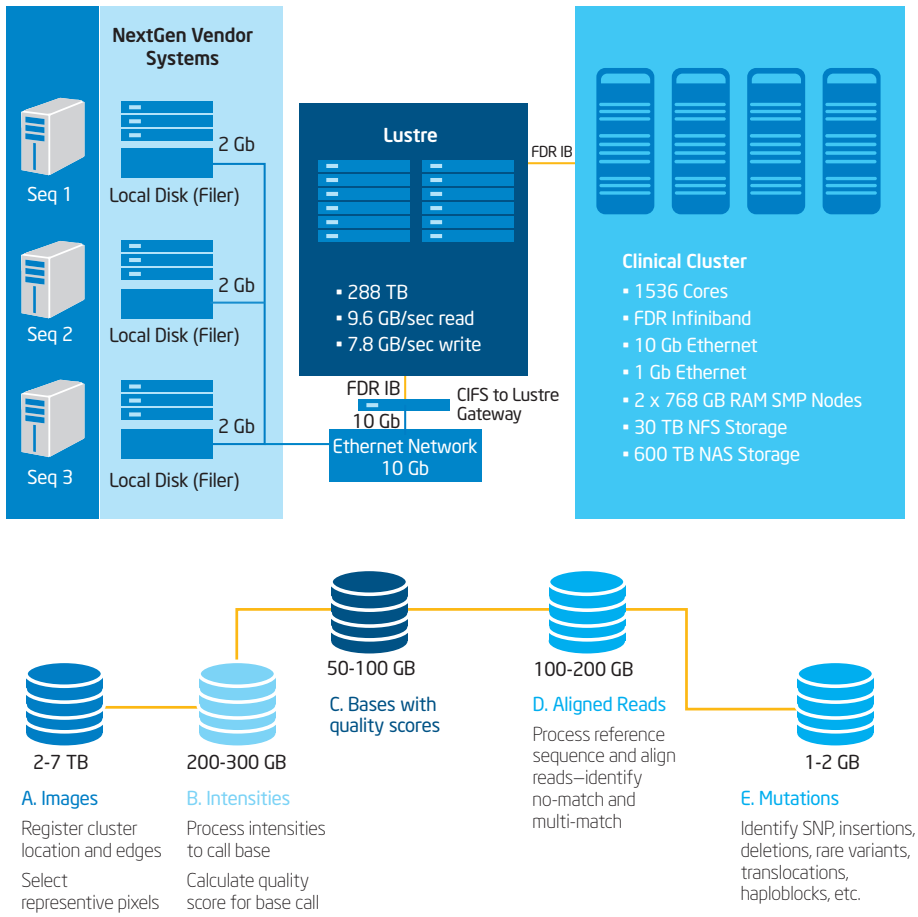

Figure 1. Next-Generation Sequencing Computing and Storage Environment.

"The TGen solution represents all of our HPC best practices and how we help our customers apply them to the genomics space," says Dr. Otero. One fruit of the collaboration is Dell's Active Infrastructure for HPC Life Sciences* platform, now available to other biomedical research and clinical organizations to help them ramp up cost-effectively and quickly.

The final TGEN Active Infrastructure for HPC Life Sciences cluster includes:

- 96 x Dell PowerEdge* M420 servers
- 1,536 Intel® Xeon® processors E5-2470 product family (2.3 GHz)
- 288 TB HSS (Lustre storage for data analysis)
- 36 TB NSS (NFS storage for /home and applications)
- Bright Cluster Manager
- Intel® Cluster Studio XE
- CIFS Gateway to HSS* (to facilitate transfer of sequencing data to HSS)
- 10 Gigabit Ethernet and InfiniBand networking
- Back-up and archive storage tiers

Figure 1 shows the TGen next-generation sequencing computing and storage environment.

Bioinformatics Acceleration Pivotal

Without doubt, the most significant performance gains resulted from accelerating bioinformatics. "We knew we could only get a 10 to 15 percent speedup relying solely on faster processors," says Dr. Otero. "Parallelizing application execution (**Bowtie 2**/**TopHat**) was necessary to get the many-fold increase we were seeking."

The particular next-gen sequencing required for the trial is called RNA-Seq, which offers several advantages over microarrays. One advantage is that RNA-Seq provides a far more precise measurement of levels of transcripts present than microarrays. Another strength is RNA-Seq's ability to identify novel mutations, not just those previously identified. Both capabilities are critical in profiling tumors (for a fuller discussion, see **RNA-Seq: A Revolutionary Tool for Transcriptomics**).[9]

Here is a glimpse of the trial tissue collection and sample process:

- Biopsies are taken at a hospital and sent (one day) to a CLIA lab where analytes (e.g., RNA or DNA) are isolated (one day).
- The CLIA lab dispenses the requested aliquot, perhaps a microgram, which is then sent (one to two days) to TGen for sequencing.
- TGen does library preparation (three to five days) and then performs a sequencing run (about five days). "Shipping to a CLIA laboratory allows us to perform research on a portion of the sample while the majority of it can remain in a CLIA chain of custody," says Dr. Huentelman.

Key steps in the RNA-Seq process and data analysis:

- RNA is extracted from the tumor and then converted to complementary DNA (cDNA), which is then sequenced in a next-gen DNA sequencer (TGen uses Illumina* instruments). The DNA sequencer output is processed on the cluster to produce a detailed list of the particular genes expressed in the tumor, including specific mutations present in those genes and how much each gene was expressed, inferred from the number of RNA transcripts present in the sample.

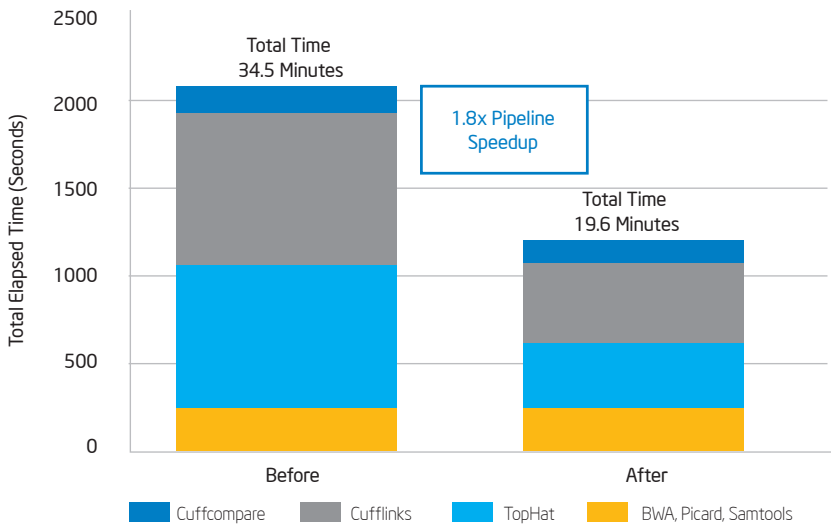

Figure 2. RNA-Seq Pipeline Improvement Over Time.[13]

- Clinicians use this data as input to the data mining platform (OncInsights*)[10] which analyzes the data and suggests best drugs for use against the specific tumor based on several factors such as proven efficacy, implicated pathways, and potential toxicities. The Tumor Board makes the final choice. This trial permitted selecting up to four drugs for use in combination therapy.

Processing RNA-Seq data can be tricky. "The difference between RNA-derived and DNA-derived reads we get off the instrument is primarily how they are handled during the alignment stage," says Dr. Huentelman. "TopHat is a software package that uses the Bowtie aligner and maps the reads to the genome. Reads that are RNA-derived can have huge gaps in their alignment to the genome because they do not contain any of the intervening DNA sequences (introns) that are removed in the cell during splicing. TopHat can address this difference and map those reads that span splice junctions and therefore might need to be split into smaller pieces that map many thousands of base pairs away from each other. In the end, one can end up with a very complete picture of the RNA (transcriptome) information content of your sample from that one snapshot in time."

Urgency to get the new system and data analysis pipeline up and running in time for the trial prompted TGen to focus mainly on accelerating alignment. "One of the biggest bottlenecks is simply mapping all of the millions of these sequence reads back to the human genome. So instead of just using one big, long file of reads and reading them line by line, the strategy is to break them into chunks and send the individual chunks to individual nodes, and then put it all back together at the end," Dr. Huentelmen says.

Bowtie 2, based on Burrows-Wheeler indexing, is the real workhorse. "Bowtie 2 is an ultra-fast and memory-efficient tool for aligning sequencing reads to long reference sequences. It is particularly good at aligning reads of about 50 up to hundreds or thousands of characters, and particularly good at aligning to relatively long (e.g., mammalian) genomes. Bowtie 2 indexes the genome with an FM index to keep its memory footprint small. For the human genome, its memory footprint is typically around 3.2 GB. Bowtie 2 supports gapped, local, and paired-end alignment modes."[11]

The Intel Xeon processor E5 family was newly available, so a fair amount of bioinformatics software wasn't yet tweaked to exploit the architecture's multithreading capability. Intel expertise and tools were crucial assets in optimizing the alignment software to do so.

Figure 2 shows the RNA-Seq pipeline improvement over time using the Intel Xeon processor E5-2687W product family (3.1 GHz).[12]

"When you run multiple threads on an application, the key metric is how much faster it runs. If you use two threads, does it run twice as fast? It's a whole lot easier to do that with just two threads than to get the same kind of scaling when you get up to 16 or 32 threads, which is available on the Intel Xeon processor E5 family," says John J. O'Neill, PhD, staff software engineer and a member of Intel's Software and Services group who worked with TGen. O'Neill reached out to Bowtie 2 author Ben Langmead of Johns Hopkins to accelerate Bowtie 2 on the Intel Xeon processor E5 family architecture. It turned out Bowtie 2 was using an older threading package that didn't scale well on these (and presumably future) processors.

Fixing data race issues is a common problem. When running multiple threads, "occasionally, each thread wants access to a shared resource, so you need to lock it to make sure that it has exclusive access to this particular piece of data," O'Neill explains. "Then you get the data, you change it, and you open it up for other threads. It just happened they were using a software package that didn't have an efficient implementation for running on the additional threads that the Intel Xeon processor E5 family architecture has. They switched to a more efficient algorithm."

O'Neill notes in this instance, the application was computationally versus I/O bound. Two Intel tools—Intel® Inspector XE (a memory and thread debugger that caught five threading errors) and Intel® VTune™ Amplifier XE (a profiler for serial and parallel performance analysis)—were effective in implementing the revised Bowtie 2 and further improving performance. This kind of optimization assistance is welcome and often needed in life science computing, says Dr. Huentelman. "Most of the tools that are built in genomics are constructed with primarily one goal in mind," he says. "Then the field will realize that they can be used for several other goals if only they can be optimized at certain steps or processes. In my opinion, optimizing an existing tool is much more attractive versus building one from scratch just to fit your use case."

Importantly, TGen will be making its RNA-Seq analysis software freely available to other researchers.[14] Dr. Trent emphasizes that part of TGen's broad mission is to help move the entire genomics research community forward, particularly where translational issues are concerned. Lessons learned here should help organizations working to create efficient genomics data analysis capabilities.

Quantitative RNA-Seq Read Counts

Accuracy, of course, matters in the clinic. A critical issue in RNA-Seq is delivering a sufficient number of counts—sequence reads matching a gene segment—to instill confidence in the result.[15]

"It's important to focus on unique sequencing counts to quantitate transcript molecules." says Dr. Huentelman. "Paired-end sequencing helps us identify the molecule more accurately because we can map it better to the genome. But those paired reads are only counting that molecule once. Our goal has always been to be somewhere around 50 million or greater counts per sample."

Not all experiments require such deep coverage, but given the nature of cancer and the associated heterogeneity of the disease it could be important to quantitate transcripts in the TGen samples. So far, TGen has been averaging 75 million counts, well above the goal and more than what's commonly reported in peer review journals.

A typical job (alignment and transcript quantification) would take 16 to 24 hours on the old system. The new TGen pipeline, optimized for parallel execution, took approximately three hours and is even faster today," says Dr. Huentelman.

"Our data is very reproducible. We don't want to guess on anything. We are getting really close clinically to having sequencing on par with some of the more advanced clinical tests, such as MRIs, on both cost and turnaround time. When you have an MRI and you get results back, how long does that take compared to how long it takes to sequence a transcriptome? This is an interesting time because the nucleic acid sequencing world is about to collide with the same sort of model that MRI uses."

Close collaboration and coordination were instrumental (and remain so) in keeping the project on track. Twice-weekly team conference calls, quarterly on-site meetings, and frequent ad hoc communication between personnel working on a particular piece of the project were all part of the process.

Face-to-face meetings were especially fruitful, says Lowey. "We also spent a lot of time on email, passing back and forth spreadsheets and Visio* drawings." Periodic meetings with the clinical team leaders, including Dr. Sholler, are also ongoing.

With the system up and running, TGen is finding more ways to drive efficiency and value. "One scientist has a LIMS system integrated with the pipeline. The bioinformatics guys have put together a system where the technician running the sequencer, just by running the jobs, initiates automated workflows on the cluster without human interaction," says Lowey. "We don't have it for every single workflow yet, but doing so wherever it makes sense is a big cost saver."

Figure 3. Dell's Active Infrastructure for HPC Life Sciences Platform.

The Intel® Xeon® Processor E5-2600 Product Family: The Heart of a Modern Data Center

Modern scientific computing is characterized by big data applications, power and space constraints, and I/O bottlenecks that drive up costs and impede throughput. Nowhere is this truer than in life sciences.

The Intel® Xeon® processor E5-2600 product family is helping systems vendors and research organizations solve these problems.

It provides more cores and cache than the previous generation, along with faster memory and additional hardware enhancements for virtualization.

The Intel Xeon processor E5-2600 product family also provides fast data movement in dense virtual environments to increase server performance and accelerate network and storage communications.

Learn more at **www.intel.com/xeon**

Future: Hadoop,* Virtualization

The first iteration of TGen's cluster has considerable short- and long-term headroom for improvement. For example, although TGen reduced its power demands significantly simply by adopting the Intel Xeon processor E5 family-based system, it has yet to fully exploit the Intel Xeon processor's power capping capability, which can individually step power up or down per rack, node, or processor.

Two more examples are Hadoop* and MPI, which were examined early in the project. The needed bioinformatics software to use those computational approaches was unavailable or deemed unready. "I think the virtualization technologies, whatever label you choose, are now sophisticated enough and rich enough that putting the next-gen workload into an environment like that makes sense," says Lowey. "We've done some testing with Hadoop and there's a particular part of the workflow for which it's really good– much faster than a traditional environment, even using a parallel C routine."

"The other piece is, I believe, as you manage clinical and payer type information that is going to flow through these systems, all of a sudden regulatory compliance becomes a big deal. One of those requirements, especially for the FDA, is repeatability. You're going to have to be able to run the same workflow seven years down the road and get the same result. As you know, at the frightening pace that IT moves, it would be extremely difficult to achieve using just raw iron, because everything is going to change. Virtualization approaches will be important here."

The same is true for software generally and bioinformatics specifically. "A year and a half is a long time," says Dr. Huentelman. "Undoubtedly, we'll be doing something very different. We might be using a new software approach and will certainly be working with new hardware. Sequencers may be different and have a different output. It's exciting, and also frustrating, because the field is constantly pushing forward and the big challenge is trying to stay up to date."

Understandably, HPC and bioinformatics tools are always works in progress. The Active Infrastructure for HPC Life Sciences* is deliberately a flexible platform and amenable to change. The collaboration with Dell and Intel will help TGen keep pace.

"There certainly will be a version of the processor based on Intel's Ivy Bridge architecture, scheduled to be available sometime in 2014," says Dr. Otero. "The other technology we are very focused on is the Intel® Xeon Phi™ coprocessor. The solution was designed with 2U of rack space that can house a Dell PowerEdge R720* server. We've got some Intel Xeon Phi coprocessors in the lab right now. TGen has one. We've been working with Intel to optimize some of the applications in TGen's pipeline so they'll compile easily on all of Intel's architectures. We are really looking at offloading elements of the pipeline onto the Intel Xeon Phi coprocessor."

A GUI interface for scientists to launch jobs is also being examined. Currently, TGen scientists are comfortable at the command line. The expectation is that as the second trial expands and there are more users—for example, Dell is assembling a similar solution for NCI to work with TGen on the project—there will be a need for an easier-to-use interface. Internally, Dell has been working with the iPlant* interface on clusters.

"When TGen gets to the point in the clinical trial where it has several investigators contributing data and trying to collaborate, something like the iPlant interface is going to be a solution." says Dr. Otero. "That's an open-source toolkit. You could have this iPlant umbrella and GUI interface for informatics and submit jobs and pick where you would want your jobs to run."

Conclusion

Modern molecular medicine is as much a tour de force of HPC as it is of advanced experimental technologies such as next-gen sequencing. Close collaboration between IT and the medical community, such as the TGen-Dell-Intel effort, is a necessary element for advancing medicine generally and personalized medicine specifically. The datasets are too large, the networks of biological interactions (e.g., disease mechanisms, drug-drug) are too complex, and the clinical need for speed is too great.

What's more dramatic is that the cost reduction in sequencing (approximately USD 3,000 now), the emergence of reliable bench-top sequencers priced well below their bigger brothers, and maturing data analysis pipelines are all adding fuel to the genomics fire.

TGen's supercomputer solution, designed for genomics analysis, is enabling Dr. Sholler to deliver personalized medicine to pediatric cancer patients today. Moreover, the lessons learned from the TGen-Dell-Intel collaboration, along with free access to the RNA-Seq data analysis pipeline software developed, will accelerate personalized medicine into the mainstream.

For more information about the technology deployed at TGen, visit **Dell's Active Infrastructure for HPC Life Sciences**.

For more information about Intel's work with the health and life sciences industry, visit **http://www.intel.com/healthcare/bigdata**.

For more information about Intel server technology, visit **www.intel.com/xeon**.

About TGen

Translational Genomics Research Institute (TGen) is a Phoenix, Arizona-based nonprofit organization dedicated to conducting groundbreaking research with life-changing results. TGen is focused on helping patients with cancer, neurological disorders, and diabetes through cutting edge translational research (the process of rapidly moving research towards patient benefit). TGen physicians and scientists work to unravel the genetic components of both common and rare complex diseases in adults and children. Working with collaborators in the scientific and medical communities literally worldwide, TGen makes a substantial contribution to help its patients through efficiency and effectiveness of the translational process. For more information, visit https://www.tgen.org.

About Dell

Dell Inc. (NASDAQ: DELL) listens to customers and delivers innovative technology and services that give them the power to do more. For more information, visit www.dell.com. Dell OEM Solutions helps its customers find more balance between execution and innovation with dedicated OEM resources, industry-standard hardware and global services and support capabilities. Dell helps its OEM customers improve their time to revenue and run their operations more efficiently for increased competitive edge. Learn more at www.dell.com/oem.

About Intel

Intel (NASDAQ: INTC) is a world leader in computing innovation. The company designs and builds the essential technologies that serve as the foundation for the world's computing devices. Learn more at newsroom.intel.com and blogs.intel.com.

[1] Giselle Sholler, MD," Molecular Guided Therapy for Refractory or Recurrent Neuroblastoma," ClinicalTrials.gov (http://clinicaltrials.gov/ct2/show/NCT01355679?term=NCT01355679&rank=1).

[2] Giselle Sholler, MD," Molecular Guided Therapy for Refractory or Recurrent Neuroblastoma," ClinicalTrials.gov (http://clinicaltrials.gov/ct2/show/NCT01355679?term=NCT01355679&rank=1). Excerpt: "Guided therapy will allow the use of any therapeutic combination (up to 4 agents) provided it includes medications contained in the study report. All patients will be treated according to the discretion of the treating oncologist and study committee (minimum 3 oncologists and one pharmacist). Extent of disease will be measured and assessed for changes throughout the course of the study and at 6-8 week intervals (every 2 cycles)."

[3] Giselle Sholler, MD, "Molecular-Guided Therapy for Relapsed and Refractory Childhood Cancer," ClinicalTrials.gov (http://clinicaltrials.gov/ct2/show/NCT01802567?term=NCT01802567&rank=1).

[4] Giselle Sholler, MD, "Progress Report NMTRC/Molecular Guided Therapy (MGT)," presented at the 2013 NMTRC Symposium at the Wyndham Lake Buena Vista Resort in Orlando, FL. (http://www.youtube.com/watch?v=IcgN2ZZc70o).

[5] Giselle Sholler, MD, "Molecular Guided Therapy for Refractory or Recurrent Neuroblastoma" (http://clinicaltrials.gov/ct2/show/NCT01355679?term=NCT01355679&ran k=1)

[6] "First Genomic-based Pediatric Trials Launched in Neuroblastoma," The ASCO Post (http://www.ascopost.com/issues/january-15-2012/first-genomic-based-pediatric-trials-launched-in-neuroblastoma.aspx)

[7] In the first trial, microarray data developed in a CLIA setting provided decisive information for the Tumor Board. TGen worked in parallel on next-generation sequencing and data analysis but in a research setting. TGen has since obtained CLIA certification, enabling use of richer next-gen data for clinical decisions in the second trial.

[8] Performance tests were conducted and are being reported, by TGen. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark* and MobileMark*, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. Test configuration: Dell PowerEdge* M420 server, Intel® Xeon® processor E5-2470 (2.30 GHz). Bios settings include: Hardware prefetch enabled, Intel® Turbo-Boost Technology disabled. For more information, contact TGen.

[9] Zhong Wang, Mark Gerstein, and Michael Snyder, " RNA-Seq: a Revolutionary Tool for Transcriptomics," NHA Public Access (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2949280/).

[10] Ken Garber, " Ready or Not: Personal Tumor Profiling Tests Take Off," British Journal of Anaesthesia (http://jnci.oxfordjournals.org/content/early/2010/12/29/jnci.djq556.full).

[11] "Bowtie 2: Fast and Sensitive Read Alignment," Johns Hopkins University (http://bowtie-bio.sourceforge.net/Bowtie 2/index.shtml).

[12] Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark* and MobileMark*, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. Configuration: Processor configuration based on current Intel benchmarks: http://www.intel.com/content/www/us/en/benchmarks/server/xeon-e5-2600-v2/xeon-e5-v2-hpc-life-sciences.html. For more information go to http://www.intel.com/performance.

[13] Test used a two-socket Intel Xeon processor E5-268W (3.1 GHz). Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Before: Bowtie2 2.0.0-beta7, TopHat 2.0.4, Cufflinks 2.0.2; After: Bowtie2 2.1.0, TopHat 2.0.8b, Cufflinks 2.1.1. Performance tests, such as SYSmark* and MobileMark*, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. Configuration: Processor configuration based on current Intel benchmarks: http://www.intel.com/content/www/us/en/benchmarks/server/xeon-e5-2600-v2/xeon-e5-v2-hpc-life-sciences.html. For more information go to http://www.intel.com/performance.

[14] TGen workflow scripts and datasets are available at http://public.tgen.org/rna-seq.

[15] "How Many Reads Are Enough?" RNA-Seq Blog (http://www.rna-seqblog.com/information/how-many-reads-are-enough/) and "A comparison of Methods for Differential Expression Analysis of RNA-Seq Data" (http://www.biomedcentral.com/1471-2105/14/91).

 0114/DW/HBD/PDF 330059-001US