



# UNLOCKING PRECISION MEDICINE WITH LOWER-COST GENOMICS

New, more affordable technologies offer researchers a deeper understanding of the genetic mutations that cause disease.

An advanced cancer patient receives two more years of life because his doctors found a clinical trial drug built for the mutation-fueled molecular pathway that drove his disease. A diagnostic odyssey comes to a halt, yielding answers for the parents of a child with a rare genetic disease. A MRSA outbreak doesn't happen at a hospital because the pediatric unit, with the help of the pathology department, finds the source and quickly contains it.

**“EVEN IN THE CASE OF NOVEL IMMUNOTHERAPY DRUGS, HAVING A BETTER UNDERSTANDING OF THE PREDICTIVE AND PROGNOSTIC BIOMARKERS FOR DOCTORS TO DETERMINE THE RIGHT COMBINATION THERAPIES FOR EACH PATIENT REQUIRES DEEPER DISCOVERY.”**

—Bryce Olson

Global Marketing Director for Health and Life Sciences, Intel

All these improved outcomes started with genomics. The ability to get to the molecular essence underpinning disease is unlocking the power of precision medicine and dramatically impacting outcomes in oncology, pediatrics, and infectious disease, among many other areas.

In the case of cancer, “precision medicine” often refers to the emerging practice of using genetic information about a patient’s tumor to diagnose or treat the disease more precisely. With this approach, physicians can select the most appropriate treatments based on their knowledge of the molecular abnormalities, such as genetic mutations, in tumors.

But what do physicians do when molecular profiling for a patient’s disease doesn’t result in clear “clinically significant mutations”?

More oncology practices are realizing that single-agent targeted therapeutic approaches don’t give patients the long-lasting, durable response rates necessary to achieve the promise of precision medicine. That’s because multiple mutations are often at work. Referred to as “tumor heterogeneity,” diseases like cancer often result in so-called “variants of unknown significance” that spawn more questions than answers.

“Even in the case of novel immunotherapy drugs, having a better understanding of the predictive and prognostic biomarkers for doctors to determine the right combination therapies for each patient requires deeper discovery,” said Bryce Olson, Intel’s global marketing director for health and life sciences.

Fortunately, advances in next-generation sequencing (NGS) technologies are resulting in much greater genome sequencing datasets for researchers. These technological advances have facilitated deeper research on the biology of things like cancer cells, and open the door to newer discoveries of potential diagnostic markers and therapeutic targets. At the same time, a new generation of clinical trials, guided by tumor profiling and genetic testing, has emerged to begin

**“AS THE DEMAND FOR GENOME SEQUENCING GROWS, SO DOES THE AMOUNT OF DATA THAT MUST BE PROCESSED, STORED, AND MANAGED.”**

—Jennifer Esposito  
Worldwide General Manager,  
Health and Life Sciences, Intel

translating basic discoveries into new diagnostic tests and targeted therapies.

“As the demand for genome sequencing grows, so does the amount of data that must be processed, stored, and managed,” said Jennifer Esposito, Intel's worldwide general manager, health and life sciences.

A high-quality whole-genome sequence is nearly 1 terabyte—about the size of 40 single-sided Blu-ray DVDs. Patients can't afford to wait weeks or months to receive a treatment plan based on analysis of that data.

“The faster the analytics can be performed, the faster a result can be delivered to determine a treatment plan,” said Dr. Michael McManus, senior health and life sciences solution architect at Intel. “Researchers want optimized solutions that they can deploy quickly and easily, that provide a more flexible and scalable solution for genomics workloads. Together with open source and commercial genomics software providers and our OEM hardware partners, we have created solutions that enable researchers to streamline next-generation sequencing workflows and significantly reduce the total cost of ownership compared to current solutions.”

In a typical genomics workflow that Dr. McManus helps clinical labs implement, DNA is extracted from a patient's blood or tumor specimen and processed by a genome sequencer, essentially digitizing the human sample. Once the sequence data is received, researchers use high-performance computing (HPC) clusters to quickly perform the genome analytics, also known as bioinformatics. The results are used to interpret any clinically significant genetic variants, and that information is used to create a treatment plan, including any specific prescriptions for the patient.

Dr. McManus, who was trained as both a polymer chemist and a synthetic organic chemist, has created guidelines for estimating the amount of computing and storage infrastructure an organization will need based on the throughput demands for different genomic workloads.

“The workflow for genomics can be viewed like a chemical reaction. We have cluster sizing guidelines for this that help labs determine what they'll need initially and what they'll need as their demand scales,” said Dr. McManus, who has been involved in computing solutions for genomics and bioinformatics for most of his career.

Using Intel's Scalable System Framework (SSF) as the basis for the reference architecture for genomics clusters, more efficient hardware clusters can be designed so fewer nodes are needed to process greater volumes of genomes. For example, QIAGEN, a bioinformatics company, runs its genomics applications on a 32-node cluster designed specifically to keep pace with a set of 10 sequencers (Illumina's HiSeq X Ten System).

“We helped QIAGEN create a solution that was capable of analyzing whole



**“FASTER ANALYTICS IS ALWAYS A PRIORITY, BUT THE NEXT BIG REQUIREMENT IS FOR RESEARCH INSTITUTIONS AND CLINICS TO RELIABLY PREDICT A GENOMICS CLUSTER'S THROUGHPUT.”**

—Dr. Michael McManus  
Senior Health and Life Sciences  
Solution Architect at Intel

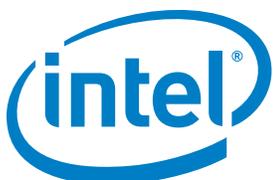
human genomes for as little as \$22 each,” Dr. McManus said. “They were able to use up to 62 percent fewer nodes than recommended by Illumina, who manufactures the industry’s highest-throughput NGS system to date, reducing their total cost of ownership of the genomics analytics solution by 47 percent<sup>1</sup>.”

“Faster analytics is always a priority, but the next big requirement is for research institutions and clinics to reliably predict a genomics cluster’s throughput,” said Dr. McManus. “In this way, they can purchase the most efficient system possible and have a metric to associate the anticipated increase in their sample volume with the required additional computing and storage hardware. This enables clinics to maximize their investment.”

Predictable throughput enables cluster scaling without guesswork. Performance is already validated so analysts can simply order more clusters if they need more capacity.

Intel has completed similar benchmarking work with the Broad Institute of MIT and Harvard, providing researchers with best-practices infrastructure for deploying the Broad’s Genome Analysis Toolkit (GATK) pipeline. The two organizations announced a \$25 million, five-year collaboration last year<sup>2</sup>, including the creation of the Intel-Broad Center for Genomic Engineering that will build, optimize, and widely share new tools and infrastructure to help scientists integrate and process genomic data.

This new effort will apply Intel’s data analytics and artificial intelligence prowess to Broad’s expertise in genomic data generation, health research, and analysis tools toward the goal of building new resources to promote biomedical discoveries, including those that advance precision medicine.



<sup>1</sup> “Analyzing Whole Human Genomes for as Little as \$22,” an Intel whitepaper

<sup>2</sup> <https://www.broadinstitute.org/news/broad-institute-teams-intel-integrate-genomic-data-diverse-sources-and-enhance-genomic-data>

© 2017 Intel Corporation. Intel, Intel Core, and the Intel logo are trademarks of the Intel Corporation or its subsidiaries in the U.S. and/or other countries.

\*Other names and brands may be claimed as the property of others.

Printed in USA 0117/DOR/PDF Please Recycle