

Scaling to the cloud to support whole-genome sequencing

DRAGEN™ platform deployed on Microsoft Azure provides ultrarapid secondary analysis capabilities in the cloud, enabling adoption of high-throughput whole-genome sequencing



PICTURED
KYLE RETTERER, MS, CHIEF
TECHNOLOGY OFFICER OF GENEDX

"We want to get to a place
where genomics isn't the end of
the journey, it's the beginning."

Introduction

GeneDx is a Maryland-based genomic analysis company that uses next-generation sequencing (NGS) for genetic testing of rare diseases. Kyle Retterer joined GeneDx in 2010 as a bioinformatics engineer, working to establish the company's infrastructure around NGS panel testing. GeneDx quickly shifted focus to building out its whole-exome sequencing (WES) capabilities, recognizing the advantages WES brought for detecting disease-associated variants. Retterer led efforts to take what they had learned from NGS targeted panels and scale to WES.

Exome sequencing continues to grow rapidly for GeneDx, making the company a market leader. Now the company is building on that expertise to introduce scalable whole-genome sequencing (WGS). "We're at a point where we can run a genome today for what it cost us to run an exome several years ago. We expect that at some point it's going to flip and become more cost effective to perform WGS than WES. We're building our data and interpretation capabilities for that future, where there will be much more territory that we can interrogate and reinterrogate over time. With WGS we will have comprehensive genome information, instead of these targeted slices"

Moving to WGS with the NovaSeq™ 6000 System

The ability increase sequencing operations has been key to GeneDx's success. Retterer recalls running targeted panels on the Illumina Genome Analyzer™ System, "We were able to run maybe a dozen samples over the course of several days. Today we can

sequence hundreds of exomes, dozens of genomes, on a single NovaSeq 6000 System. We're much higher throughput. We're able to investigate more of the genome for more people at a lower cost because of Illumina sequencing technology." But scaling up their sequencing capabilities meant they also needed to increase their ability to manage, analyze, and interpret all that data.

"DRAGEN on Azure enables people to share data much more readily because the data are already sitting in a secure cloud...Also, the ability to reanalyze older data becomes more practical...With cloud-based storage, you have all that data readily available. It enables you to do more retrospective analyses."

"You've got to evolve the technology on all sides. The laboratory technology and the data technology are complimentary pieces."

Enter the DRAGEN Bio-IT Platform

The DRAGEN Bio-IT Platform uses reconfigurable field-programmable gate array (FPGA) technology to deliver highly accurate, ultrarapid secondary analysis of NGS data. GeneDx integrated the DRAGEN Server into its existing on-premise solution to lower costs and drastically reduce analysis times. The DRAGEN platform provided more than just the speed. "We liked the simplicity of it. We also liked the modularity of it. We were able to take DRAGEN software and plug it into our existing infrastructure and pipelines relatively seamlessly."

By incorporating a [DRAGEN on-premise solution](#), GeneDx was able to build up a significant data center with the ability to handle the computation and data storage needs that running ~5000-10,000 exomes per month required. But as GeneDx began the process of moving from WES to WGS, the challenge became obvious to Retterer. "With WGS, you generate about 25x as much sequencing data per sample as compared to WES. We've run over 300,000 exomes to date, about a third of those in the last year. But if we convert all of our WES to WGS, suddenly our data center has to grow 20-30 times its current size. The amount of infrastructure we would need, it's a bit overwhelming. Our data center would become bigger than our laboratory. We want to keep with the GeneDx mission; we're a research and diagnostics company, not a data center."

Scaling data infrastructure to the cloud with Microsoft Azure

As GeneDx transitioned from exomes to genomes, Retterer knew their use of the DRAGEN platform had to scale as well. "You've got to evolve the technology on all sides. The laboratory technology and the data technology are complementary pieces. I think cloud-based computing and storage are necessary as you adopt WGS at the scale GeneDx is doing." DRAGEN on Azure provided GeneDx the cloud-based computing they needed to scale to WGS, while delivering the same speed and simplicity they experienced with DRAGEN on-premise for data analysis.

In addition to the benefits seen with performing data analysis in the cloud, Retterer recognizes other impacts DRAGEN on Azure has on their data management. "One thing that's clear is DRAGEN on Azure enables people to share data much more readily because the data are already sitting in a secure cloud. Sharing data with clients or collaborators becomes a bit smoother. Also, the ability to reanalyze older data becomes more practical. We've sequenced 300,000 exomes, but we don't have all 300,000 exomes constantly available on high-performance computer systems on premise because that's too expensive. With cloud-based storage, you can have all that data more readily available. It enables you to do more retrospective analyses."

As a bonus, it turned out that migrating to the cloud to leverage the FPGAs on Azure was more cost efficient for GeneDx than maintaining their own high-performance computing center. "The DRAGEN platform uses these FPGA systems for analysis, and there are plenty of them in Azure. We're able to bring a large number of them online at once, run all of our analysis jobs in parallel, and then bring them offline. Because the DRAGEN platform is so fast, we don't need them for that long, so we can be cost effective with our cloud usage." That ability to analyze multiple samples simultaneously is another advantage of DRAGEN on Azure. Cloud-based analysis reduces turnaround times by eliminating the need to queue up multiple samples for analysis on a single DRAGEN Server on premise.

Security in the cloud

As a research and diagnostics company dealing with personal genomic data, security is paramount to GeneDx. According to Retterer. "Cloud-based data management can be more secure than storage on premise, because you have this larger ecosystem of security that you can take advantage of, but make sure you design for that from the start. We've worked with a managed cloud service provider, who specializes in the healthcare space, to help us with the transition and with securing all of the infrastructure. Whether you choose to work with a managed cloud service provider or use services directly through Illumina, it makes sense to get help."

Retterer sees another area where moving to the cloud offers clear advantages. "Transitioning to Microsoft Azure removes some of the risk involved in having a large, on-premise data center. We have built-in redundancy and a disaster recovery plan, but with DRAGEN on Azure, an on-premise disaster won't impact operations to the same extent."

"We liked the simplicity of it. We also liked the modularity of it... we were able to take DRAGEN software and plug it into our existing infrastructure and pipelines relatively seamlessly."

"The biggest thing is that DRAGEN on Azure enables the shift from exomes to whole genomes and the complimentary shift to broader testing of more people."

Vision for the future

GeneDx is working to integrate genomics into routine health care to improve health for everyone. Historically, GeneDx has become involved only after a genetics-savvy physician has ordered a genomic analysis, often at the end of a rare disease diagnostic odyssey. The challenge GeneDx is taking on, as Retterer puts it, is “how do you make genomics something that becomes a routine part of the healthcare journey for every kid with a developmental problem or anyone else who could benefit? How do we get them sequenced earlier, so we can get answers back sooner?”

The answer, in part, is a matter of scale. As Retterer says, “The biggest thing is that DRAGEN on Azure enables the shift from exomes to whole genomes, and the complementary shift to broader testing of more people. It would have been much more challenging to accomplish that shift from a capital and operational perspective if we were still using an on-premise data center.” Looking to the future, Retterer believes WGS is key to helping GeneDx achieve its mission. “We want to get to a place where genomics isn’t the end of the journey, it’s the beginning. We want to shortcut the diagnostic odyssey and give people more comprehensive answers much earlier in the process, enabling possible elimination or mitigation of various rare diseases through emerging therapies.”

Learn more

DRAGEN Bio-IT Platform, illumina.com/products/by-type/informatics-products/dragen-bio-it-platform

DRAGEN on Azure, illumina.github.io/dragen-azure-quick-start/3.9.5/

GeneDx, genedx.com

illumina[®]

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel | techsupport@illumina.com | www.illumina.com

© 2021 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners.

For specific trademark information, see www.illumina.com/company/legal.html.

M-GL-00416 v1.0