Efficient cloud data analysis for COPD multiomics project

Researchers at Okayama University in Japan use Illumina Connected Analytics with DRAGEN™ pipelines for analyzing whole-genome, exome, transcriptome, and metagenome data

Shuta Tomida, PhD
Associate Professor
Center for Comprehensive
Genomics Medicine,
Okayama University Hospital

"Being able to run the DRAGEN pipelines in a cloud environment via Illumina Connected Analytics was a great advantage."

Shuta Tomida, PhD leads a bioinformatics team focused on precision cancer research at Okayama University in Japan. His team had five months to complete a multiomics project investigating chronic obstructive pulmonary disease (COPD), a known risk factor for lung cancer. They used 50 samples each for germline whole-genome sequencing (WGS) from blood, tumor whole-exome sequencing (WES) from lung tissue, whole-transcriptome RNA sequencing (RNA-Seq) from lung tissue, and shotgun metagenomics from lung tissue. Dr Tomida partnered with Illumina to optimize his next-generation sequencing (NGS) bioinformatics pipelines to meet his tight deadline.

Secure, efficient secondary analysis

"One of the most challenging tasks was completing the analysis of whole-genome sequencing on time," said Dr Tomida. For this project, the Okayama University researchers were required to use a data analysis solution housed in Japan. The team decided to access cloud-based Illumina Connected Analytics, available on the Amazon Web Services (AWS) Tokyo server.

"As a prerequisite, we were looking for an environment where we could use high-spec computing resources that are difficult to build and maintain onsite," said Dr. Tomida. "We selected Illumina Connected Analytics, which has various ready-to-use analysis pipelines, such as DRAGEN Germline for WGS, DRAGEN Enrichment for WES, and DRAGEN RNA for RNA-Seq. In addition, the data analysis workflow using the cloud was very easy to use. Even bioinformaticians who are not physician scientists can use Connected Analytics to analyze NGS data."

For Research Use Only. Not for use in diagnostic procedures.

M-GL-01639 v1.0 | 1

Collaboration for custom pipelines

"We were satisfied with the Illumina team overall, including the onboarding and training program," said Dr Tomida. "The Illumina support team collaborated with us and responded quickly to resolve any issues, allowing us to complete our original plan on time."

To complete the metagenomic analysis component, Dr Tomida and his team used a custom DRAGEN Metagenomics pipeline built in collaboration with the Illumina team. The pipeline was deployed on the Connected Analytics Japan instance to enable Dr Tomida to perform this portion of the analysis.

Scalability in the cloud

"Overall, we were able to complete the data upload, analysis, and download within the initial deadline. The cost of these analyses was also within the original plan," said Dr Tomida.

"What we learned from our experience using Connected Analytics is that the cloud environment is wonderful. In our Connected Analytics environment, the WES analysis time for 50–60 Gb input was about 50 minutes, and the WGS analysis time for 130–160 Gb input was about 2.5 hours. It is difficult to build and maintain such a computing environment onsite, and being able to run the DRAGEN pipelines in a cloud environment via Connected Analytics was a great advantage."

Dr Tomida has been successful at obtaining funding for increasing the sample size for a follow-up multiomics project and plans to continue using Connected Analytics. He predicts that with Illumina Connected Analytics and DRAGEN analysis, "NGS data analysis will no longer be the rate limiting step of research."

Learn more

Illumina Connected Analytics

illumına®

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