

Review: How 3 Life Science Organizations Harness the Power of the Cloud

Cathy Ford

Whether you are a small biotech startup or a multi-billion dollar pharma giant, those that use the cloud to comb through vast amounts of information are certainly better positioned to do more with their data. We've seen headlines examples of how the cloud has helped organizations in fintech, manufacturing, and retail, but what about life sciences?

I wanted to showcase some interesting reports about how organizations in the life sciences realm are harnessing the capabilities of the cloud to uncover insights lying deep in their data to solve the world's problems in human disease and beyond.



Why Cloud?

The cloud is touted as a seemingly limitless space for compute and storage, making it a perfect tool to scale up analyses or storage for petabytes of data. In Microsoft Azure, the creation of a genomics data lake as a central repository to store all sorts of multi-omic data has shown to be an excellent tool as it allows for the connection of scalable compute services like Azure Databricks and Azure Machine Learning. Plus, the cloud comes with security capabilities that are much easier to manage than doing so with on-premise architectures.



Read my recent genomics eBook: [Building a Genomics Data Lake in Azure](#).

Moderna

Last year, COVID-19 very quickly made Moderna a household name. Starting in 2010, the company was among the first to use messenger RNA (mRNA) as a therapeutic tool. While this type of therapy was initially used for the treatment of cancers and other human diseases, once an appropriate cellular delivery mechanism (the lipid-based nanoparticle system) was hammered out, this opened the door to all sorts of therapeutic opportunities.

moderna

For SARS-CoV-2, they were able to get their vaccine candidate ready for Phase I trial just 42 days after the sequence of the Wuhan, China reference isolate became available. How, you ask? The cloud.

For Moderna, the trick is being able to test various mRNA sequences and simulate their interactions with the human body. This often relies on a complex, computationally-intensive pipeline. So, the ability to spin up this pipeline to analyze a new mRNA candidate, scale out the compute to whatever size is needed, and then spin everything back down once the results are in is key to a performant computational biology workflow.



Read more about Moderna's use of the cloud in [Business View](#).

Regeneron

Regeneron is an interesting pharma company because it has basically always operated as a data-driven biotech organization. They focus on creating monoclonal antibody treatments for everything from autoinflammatory diseases (e.g., dermatitis) to hypercholesterolemia, to cancer. The discovery and creation of these highly specific proteins (antibodies) requires the analysis of tons of genomic data, protein structure modeling, and clinical understanding.



Regeneron has a world-class genomics data center that is focused on data scalability. For them, the ability to take raw data (such as electronic health records, genome sequences, etc.) and turn it into usable information is key to designing treatments. They use Databricks to perform analyses at scale using distributed cloud computing. In fact, Databricks + Regeneron formed the collaborative research project called [Project Glimpse](#), which is a library for large-scale genomic analysis using Apache Spark.

This cloud-centered genomics practice enables massive scale data engineering and collaborative data science on billions of data points at Regeneron.

Read more about Regeneron's data-focused genetics center in [TechCrunch](#).

Icahn School of Medicine at Mount Sinai



Icahn School of Medicine at Mount Sinai

Another research group that's utilizing the cloud to harness its power for a patient. In the Klein Lab at Mt. Sinai, they use clinical and genomic datasets for gene variants in cancer risks and they use the Azure cloud to analyze large datasets.

Genome-wide association studies (GWAS) use a ton of data, but are really useful in comparing genetic information from healthy and non-healthy individuals. So, using performance pipelines in the cloud is key to doing GWAS analysis at scale.

Specifically, the Klein Lab is using the [Microsoft Genomics](#) service, which is a GATK-compliant pipeline for aligning and variant calling human sequence data. This service is a cost-effective and efficient implementation of the Burrows-Wheeler Aligner (BWA) and the GATK HaplotypeCaller pipelines.

Read more about the use of Microsoft Genomics at Mt. Sinai [here](#).

How can BlueGranite Help?

If you're reading this post, you're probably no stranger to BlueGranite. We have industry experts with experience in both healthcare and life sciences, plus a deep expertise in all things cloud.

Also, we recently created a Genomics-on-the-page to highlight solutions and technologies that are specific to genomics. Check it out at [BlueGranite.com/Genomics](#).