

Recursion Pharma Targets 100 Genetic Diseases With Splunk and Machine Learning



RECURSION
pharmaceuticals

Executive summary

Salt Lake City-based Recursion Pharmaceuticals set a goal: discover new treatments for 100 genetic diseases by 2025 using an innovative combination of biology, automation and machine learning. Today Recursion uses robotic microscopes to capture tens of thousands of cell culture images, generating terabytes of data daily. As a rapidly growing startup with just a few employees, it faced challenges in logging procedures and tracking operational data. Since deploying Splunk Enterprise, the company has seen benefits including:

- Time to value in three days, with complex laboratory adoption complete in only three months
- Visibility into all parts of the production process informs disease-targeting experiments

Why Splunk

In the past, Recursion found it difficult to manage large amounts of time-series data collected from computer-controlled instruments and video footage generated from cameras in the laboratory. The initial data management strategy hardly matched the firm's aggressive high-volume ambitions—its laboratory's microscopes currently produce on the order of 700,000 TIFF files each week, representing an 800 percent increase in productivity over 10 months.

While the company considered open-source alternatives, Ben Miller, director of high-throughput science (HTS) operations, saw the pivotal role that Splunk Enterprise could fill as Recursion ramped up its capabilities. "I was getting value out of Splunk Enterprise within about three days," Miller says.

Recursion has built a world-class proprietary machine learning system that analyzes terabytes of experimental image data daily to discover new treatments for critical diseases. This system is integrated with Splunk Enterprise via the Splunk SDK for Python, which passes operational data back into the experiment-design processes, and Splunk DB Connect to enrich log data with quality metrics. From there the Splunk Machine Learning Toolkit makes it easy to comb these higher-level operational metrics for new insights into laboratory processes—in Miller's words, to "wrangle really large quantities of data and understand what correlations are happening as they are happening, not months later."

Industry

- Healthcare

Splunk Use Cases

- Business Analytics
- Internet of Things

Challenges

- Rocket-fast ramp-up and massive data volume requires extreme scalability
- Aggressive target of 100 new treatments calls for real-time visibility into experiments
- Pulling together critical information from different sources and test instruments
- Diverse users and requirements, from Python-programming scientists to dashboard-browsing execs and techs

Business Impact

- Initial time to value in only three days, with complex laboratory ramping up operations in only three months
- Massive scale while managing data volume of 700,000 TIFF files weekly
- Timely visibility into all parts of the production facility informs experiments
- Robust platform for central log management and review
- Machine Learning Toolkit provides otherwise hidden insights into experiments
- Flexible data management platform for moving information in and out of data scientists' code base
- Substantial cost savings versus specialized laboratory information management software

Data Sources

- Log files from scheduling computers (systems that govern all computer-controlled instruments)
- Log files from individual instruments such as:
 - Acoustic droplet dispensers (devices that add reagents to cell plates)
 - Plate fluorimeters
 - Robotic microscopes

Splunk Products

- Splunk Enterprise
- Splunk DB Connect
- Splunk Machine Learning Toolkit
- Splunk SDK for Python

Using artificial intelligence (AI) and bioinformatics

In a sense, Recursion's approach is parallel rather than serial. Instead of studying an explicit molecular target related to a specific disease, the company generates and analyzes thousands of cell cultures in the presence of siRNA—a compound that “turns off” specifically selected genes creating a model of genetic disease—and drug compounds daily, using a deep-learning algorithm to process the results and AI to help determine which drugs merit further study.

Implemented in a mere three months, the Splunk platform plays two critical roles in running and improving such a complex laboratory. Splunk Enterprise helps monitor and diagnose issues in real time with complex lab instruments, catching anomalies in automated operations, letting the high-throughput science team build dashboards to measure quality over time. It also serves as a data management platform that feeds machine data back to the data scientists who work with Splunk add-ons such as Splunk DB Connect, enabling the team to share discovered knowledge.

Keeping track of the Internet of Things

The company's three large automated work cells are built around robotic arms that move plates of cells and reagents from one instrument to another. Each work cell communicates over its own subnet, with each instrument feeding logs and other data to the Splunk platform, so the team can gather analytics—and gain insights into how the complex process is working.

For example, Recursion's acoustic dispensers take thousands of measurements of drug and siRNA source plate volumes every hour, allowing for careful tracking of these valuable reagents. Recursion's scientists can track evaporation and hydration rates of these source wells to more effectively control their concentration during

“Splunk Enterprise gives us that visibility into all parts of the production facility in a very timely fashion to help inform our next set of experiments.”

John Pereira,
Chief Operating Officer, Recursion Pharmaceuticals

“The scale and speed with which we work now would be impossible to do manually. As far as capturing the data and what we do with it, it just wouldn't be possible.”

Ben Miller,
Director of HTS Operations, Recursion Pharmaceuticals

experimentation and reduce waste. In addition, they can more efficiently plan and execute experiments without being surprised by a four- to six-week wait and a \$5,000 charge for a new plate prepared by an outside manufacturer.

Making the business plan possible

Asked to estimate how much time the company has saved by implementing Splunk Enterprise, Miller says, “If you combine the amount of data we are collecting with the speed and scale at which we work, it would be impossible to do it all manually. Splunk is more than just a resource for our team, it's a requirement.”

Splunk dashboards have become the first things the HTS technicians look at before starting the day's work, to make sure there is enough source plate volume to proceed and, after each run, to confirm that the tens of thousands of planned drug and siRNA transfers were executed properly. Miller describes one microscope workstation that came with its own software. It was faster to look at the log information via the Splunk interface than learning to use the software to view the log data natively.

Looking ahead, Recursion anticipates Splunk solutions will help with everything from planning for future expansion and purchase decisions to maintaining the company's breakneck pace toward its goal of 100 new treatments by 2025.

“In the biotech industry, there is a software category called LIMS, laboratory information management software, and all of it is expensive and inflexible in comparison to Splunk. I think biotech and pharmaceutical companies should strongly consider Splunk when looking for an external LIMS system. This is the second biotech company where I have implemented it and it pays off.”

Ben Miller,
Director of HTS Operations, Recursion Pharmaceuticals

Download Splunk for free or get started with the free cloud trial. Whether cloud, on-premises, or for large or small teams, Splunk has a deployment model that will fit your needs.



Learn more: www.splunk.com/asksales

www.splunk.com