

# Pediatric brain research laboratory uses Globus Genomics to overcome IT hurdles



William Dobyns is a Professor of Pediatrics and Neurology at the University of Washington and is a principal investigator at the Center for Integrative Brain Research at Seattle Children's Research Institute. Dr. Dobyns leads a research program investigating the nature and causes of a wide range of developmental brain disorders in humans. His research in this area has led to major discoveries in the field of human genetics, particularly the nature and genetic causes of neurodevelopmental disorders.

The Dobyns research team conducts their research by collecting tissue samples from young patients and their families and then extracting, sequencing, and analyzing the genetic material in pursuit of the underlying cause. The lab utilizes next generation sequencing (NGS) methods such as whole exome sequencing. NGS is performed at any of three different sequencing centers: the University of Washington's on-campus core genomic facility, PerkinElmer, and the Broad Institute. Once sequenced, data are returned to the research team, analyzed, stored and shared with collaborators. NGS generates large data sets, the analysis of which is typically time-consuming and uses considerable computing power.

## Why is this difficult?

Dobyns and his team encountered three significant IT-related challenges: data transfer, storage and sharing, getting sufficient computing power for their analyses, and maintaining technical expertise despite normal research staff turnover.

**Moving and sharing NGS data is cumbersome.** The amount of data generated by NGS is quite large—measured in terabytes. The current practice at some sequencing centers is to physically ship hard drives loaded with the data to their customers, or upload the data onto local, but temporary, servers. When using external hard drives, the drives can't be delivered until they're fully loaded with all of the customer's data, and analysis can't begin until the hard drives arrive at the lab. Furthermore, data are only delivered to a single location, while researchers may be at several locations, all of which ultimately need access to the data. Therefore, data transfer, storage and sharing are an overall time- and effort-consuming processes.



Learn more at [globus.org/genomics](http://globus.org/genomics)

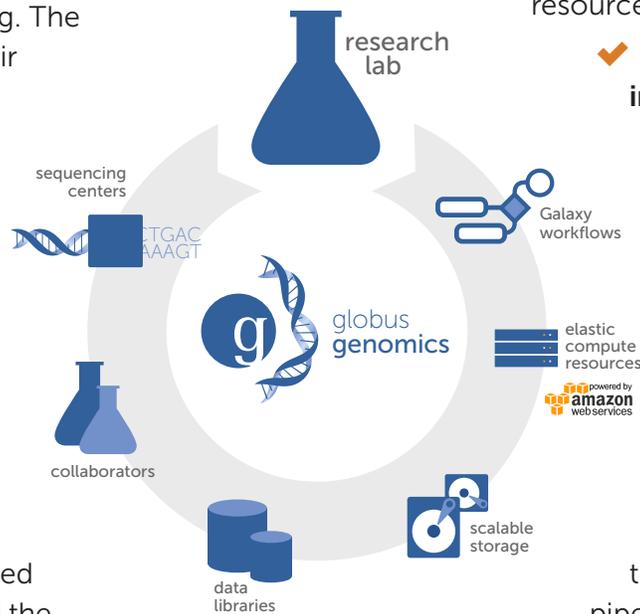
**Accessing computing power is expensive and inflexible.** “Our local storage is at or near capacity,” says Dobyns. “Our local computing environment was already being utilized heavily, and we needed a way to flexibly grow IT resources.” Analyzing a single sample takes between 12 and 24 hours of computation, depending on the size of the genome, the depth of coverage, and other variables. The team ideally wanted a pay-as-you-go model, where the team could essentially have infinite computing power available on demand, without having to maintain significant computational infrastructure.

**Maintaining technical expertise is difficult and time-consuming.** The team utilizes complicated analysis methods, requiring significant expertise to keep them running. The team would rather expend their time and mental resources interpreting results than maintaining infrastructure and applications. But their analysis tools are typically run on the command line using custom scripts that the group itself has developed, so the team has to maintain the scripts and make sure they run properly. As new members join the team, bringing them up to speed on how the analysis tools (and the systems around them) work has become burdensome. The complexity has made it critical that the team test and validate its own work extensively so that every detail of the analysis is ultimately understood by the researchers. The intricacy of the analysis pipelines also made it time-consuming to assess the potential value of new tools.

## Using Globus Genomics

Dr. Dobyns needed a solution that would allow his team to focus on their research, as opposed to solving IT problems. Globus Genomics was able to offer solutions to three critical challenges in the Dobyns team’s work.

- ✔ **Fast, reliable, easy-to-use data movement capabilities.** Both PerkinElmer and Broad Institute provide Globus Online “endpoints” for electronic data transfer. The research team was able to use Globus Online to obtain data from endpoints with very little effort and in a time-efficient way, despite the huge size of data involved.
- ✔ **A cost-effective, scalable, cloud-based computing system.** Globus Genomics uses Amazon Web Services (AWS) for computational analysis. AWS makes it easy for Globus Genomics to adjust the resources it uses in response to changing research requirements, providing a near-infinite pool of available resources, while only charging for the actual resources used.



- ✔ **A flexible, transparent, intuitive environment for defining and executing analysis pipelines.** The Galaxy-based workflow engine in Globus Genomics made it easy for the researchers to capture the steps in their current pipelines, express them in graphical form for easy comprehension and learning throughout the team, and to execute the pipelines without losing the ability to “tweak” parameters.

## Results

The results have been very good for the Dobyns lab.

The team’s days of wrangling hard drives via FedEx are over. By using Globus Online to transfer their data, they are able to move and share tens of terabytes of data quickly and reliably. Instead of taking weeks to months for the data to be available for analysis, it is now available within hours. “Globus Genomics is a very powerful platform,” says Dobyns’ collaborator, Dr. Ghayda Mirzaa, M.D., at the Center for Integrative Brain Research

and Seattle Children's Research Institute, "and the integrated data movement capability has become one of our favorite tools for next-generation data analysis, allowing us to transfer large data sets easily and efficiently."

Globus Genomics has allowed the team to analyze data at least five times faster. The elastic computing environment can execute analysis pipelines in parallel (they have run ten pipelines at a time successfully), which accelerates time-to-results and ultimately, time-to-science.



The Dobyns and Millen lab.

The team's analysis pipelines are encoded in the system, which means the team can continue to use pipelines that were established by former team members. The graphical environment allows newer team members to easily learn how the pipelines work and to "tweak" the pipelines to use newer methods or tools. The environment makes it easy to adjust or extend the pipelines for comparative analysis and further downstream interpretation.

"Globus Genomics has been extremely effective for our exome analysis pipelines," says Dr. Dobyns. "The platform is easy to use and running in a cloud-based environment has relieved our dependence on local server availability and local storage space, which is especially important as our analysis needs continue to grow."

With Globus Genomics, the Dobyns lab now plans to increase the scale of their analysis. They are exploring the genomic interpretations that Globus Genomics offers and they may also use the platform for data archiving and enhanced data sharing.

Contact us to learn more about how Globus Genomics could enable research and simplify data management in your lab.

Email us at [genomics@globus.org](mailto:genomics@globus.org)