Harnessing genomic data to accelerate precision medicine

SCALE WITH THE VELOCITY OF TECHNOLOGICAL INNOVATION

DNAnexus® | Microsoft Azure

The promise of precision medicine and better patient outcomes is here, through the utilization and application of genomic data in biomarker discovery and clinical research. Breakthroughs in preventive care, treatment of illness and disorders, and new medicines are happening every day in the world's most advanced research and healthcare institutions, fueled by genomic science.

And yet, the surface has only been scratched. This is still a relatively new field and despite the progress to date, the potential still seems boundless.

Information technology plays a key role. Data-handling and processing capabilities will need to continue to scale beyond what the largest and most capable healthcare IT organizations have seen. In fact, a single human genome takes up to 100 GB of storage space.

Because genomics is an emerging field, the IT used to support it often lacks maturity. For example, an organization's existing IT might include:

- On-premises datacenters
- Homegrown, do-it-yourself (DIY) infrastructure
- Limited data storage and handling capabilities
- Inadequate security and compliance protections
- Firewalls that make collaboration difficult—with outside institutions and even other geographical locations within the same organization

> "As data sets get larger, there are no limits to the number of biological insights that can be found, but it has also been clear that we need novel and improved methods to help us analyze these data sets."
>
> BARBARA CHEIFET
> "Where Is Genomics Going Next?," *Genome Biology*

Genomic research is data intensive, to say the least, and the need for larger and more complex genomic data sets is becoming a reality for driving innovation. Including genomic data in clinical and drug discovery pipelines adds an additional layer of complexity and scale in working with clinical and phenotypic data. Researchers need a solution that not only handles the increasing volume requirements but one that also lets them easily query and analyze all data with ease.

While technology-enabled genomic data science moves quickly from research to the clinic, IT maturity must also rise to meet its new demands in order to capitalize on the opportunity and move healthcare forward.

In this e-book, we'll lay the groundwork for the state of genomics and provide recommendations for how to modernize your IT through advanced cloud-based systems. We'll show how DNAnexus and Microsoft have tackled these challenges to create solutions that help their customers advance precision medicine.

"Researchers believe that cloud computing is the only storage model that can provide the elastic scale needed for DNA sequencing."

"The Genomic Data Challenges of the Future," *The Medical Futurist*

The state of genomic exploration and application

Among researchers in the industry, there is tremendous momentum around this topic.

According to a study by Tufts University,[1] pharmaceutical companies have nearly doubled their R&D investment in personalized medicines over the past five years and expect to increase it by an additional 33% in the coming five years. Some 42% of all drugs in development today are "personalized medicines," and that number jumps to more than 70% when applied to oncology drugs alone.

Even with recent advancements in the field of genomics, there are still massive swaths of greenfield to stake out and capture—ground that your competitors are eyeing as you read this.

The growth potential for your organization is completely worth investigation and investment, and your ability to positively impact patient outcomes—the best reason to enhance your capabilities and services—could be transformative.

> "Researchers estimate that between 100 million and as many as 2 billion human genomes could be sequenced by 2025."
>
> "The Genomic Data Challenges of the Future," *The Medical Futurist*

KEY FACTORS

Let's explore the key factors that are shaping today's genomics marketplace.

Simple economics

Because of technological leaps, the price of next-generation sequencing (NGS) has fallen to the point where using the data in a variety of research applications has become dramatically more accessible. As costs drop, demand rises, creating an unprecedented need to scale, thus transforming data uses, architectures, and service delivery models.

"The cost of sequencing a human genome has dropped from $2.7 billion in 2003 to approximately $1,500 in 2018. In addition, an increasing number of people are using direct-to-consumer genetic tests that analyze partial yet interesting genome regions at a cost of less than $100. In the coming era of precision medicine, more and more people will be obtaining their own whole-genome sequence."

SANG YUP LEE
Korea Advanced Institute of Science and Technology

This transformation, in turn, raises the expectations of researchers, who constantly crave faster performance, easier self-service, and greater computational reproducibility—qualities DIY solutions rarely deliver. This virtuous cycle continues as NGS technology keeps improving, prices keep falling, demand keeps rising, and exponentially more—and more meaningful—data is produced.

Data islands

Collaboration is the backbone of scientific discovery. Researchers need to work together easily while mitigating potential violations in security and compliance. Large silos of data severely hinder this. Even within the same organization, researchers may struggle to share insights without a modernized infrastructure.

"Cloud computing is a popular and enduring paradigm that has changed how large and small companies manage computational resources. Increasingly, it is also changing how scientists in genomics and in other fields collaborate and deal with vast archive data sets."

BEN LANGMEAD AND ABHINAV NELLORE
"Cloud Computing as a Platform for Genomic Data Analysis and Collaboration,"
National Institutes of Health

Different teams may be unable to work on a project simultaneously, having to wait until the previous work has been done before doing their part in isolation.

Without the ability to leverage cloud-based and secure collaboration methods, organizations have to share NGS data with others by shipping hard drives back and forth. This drastically slows progress and limits opportunities for cooperative breakthroughs.

Fear of change

Internal stakeholders may be dedicated to the homegrown solution they know and may defend it as specific to their unique needs—even when a better solution may be available. Many researchers worry that third-party infrastructure won't give them the flexibility to analyze data for their specific research, believing their workflows, tools, and apps couldn't be leveraged when migrated to turn-key and out-of-the-box cloud-based solutions.

The transition seems daunting, but the cost of doing nothing may be a higher risk. Organizations wasting considerable resources, both time and financial, to maintain their homegrown solutions find themselves not focusing their efforts on the purpose of scientific discovery. Instead of being active participants in the velocity of innovation, they become mere observers.

What's more, when a researcher leaves the organization, a new employee must learn the intricacies of the DIY solution—how it was built, how to maintain it, and how to operate it. This prevents teams from hitting the ground running and slows down the entire process of discovery.

Security inexperience

Complexities around managing privacy, security, and compliance regulations globally and regionally are a critical component of ensuring researchers do not put their organizations at risk.

Regulations such as ISO 27001, HIPAA, CAP/CLIA, GxP, and GDPR are only the tip of the iceberg. Researchers are tasked with becoming experts on all regulations at a global scale, regulations that are continuously and rapidly changing.

A comprehensive framework for security, quality, and privacy must be established and upheld in addition to layers of features allowing for auditability, trackability, and reproducibility.

Organizations need to easily access controls, allowing for granular privileges to be specified on groups of users and data and complying with localization requirements, and all confidential data should be protected against unauthorized access through two-factor authentication, password complexity and change requirements, and session time-out features. Additionally, researchers will need to provide detailed audit trails necessary to document compliance.

Informatics solutions need to address a broad range of security and compliance requirements, enabling researchers and bioinformaticians to focus more on their scientific breakthroughs.

Questions to ask when evaluating your current informatics solution

Can my current infrastructure handle the scale and scope of tomorrow? Six months from now? Two years from now?

Industry trends confirm more genomic data will be incorporated into clinical and drug discovery pipelines, and the complexities will only increase. A paper published in the journal *PLOS Biology* estimated that "as much as 40 exabytes of storage capacity will be needed by 2025 just for human genomes."[2]

Many homegrown solutions are tasked with solving the challenges of today, without any foresight into the challenges of tomorrow. Systems originally designed for lower volumes of similar data will face difficulty deploying new pipelines and workflows, and will experience challenges finding and accessing data easily.

Was my DIY system built to meet all regional and global security and compliance requirements without exposing us to risk?

There are three critical dimensions involved in examining a genomic informatics platform: security, privacy, and quality. Each serves as a vital component in truly sustainable science. All three depend on robust authentication and authorization, track-and-trace functionality of all data, and clear-cut governance. DIY systems increase the burden on researchers, as they must be self-reliant in their security and compliance expertise.

"One grand challenge of genomics research in the next decade is to enable functional biology at scale, finally making the mechanistic dissection of biological processes a high-throughput endeavor and overcoming the 'one gene, one postdoc' paradigm of molecular biology."

CHRISTOPH BOCK
CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences

Does my system make it easy to work with collaborators across the hall and around the globe without exposing my organization to potential risks?

The task of improving patient health requires the scientific community to collaborate simultaneously on innovation. Collaboration can be internally within an organization or between multiple organizations across multiple sites throughout the world.

Your system needs fine-grained access controls paired with a global security and compliance framework that allows seamless collaboration, all while protecting sensitive data.

> "By making it easier to leverage public data, cloud computing encourages another dimension of 'strength borrowing.' That is, researchers can use public data to boost the power available to analyze a locally-generated data set, a paradigm that already prevails in microarray data analysis."
>
> BEN LANGMEAD AND ABHINAV NELLORE
> "Cloud Computing as a Platform for Genomic Data Analysis and Collaboration,"
> National Institutes of Health

Does my current cloud-based analysis infrastructure give us best-in-class and industry-standard genomic analysis workflows and let us import our own tools?

Many pre-built genomic informatics solutions offer integration but are often too rigid in their analysis options. Researchers with existing analysis pipelines will eventually require the flexibility to port them to a cloud system. They can also benefit from leveraging preloaded best-practice workflows to accelerate scientific discoveries.

Is my current infrastructure letting me use my resources efficiently, and do I need to worry about maintenance to keep up with the velocity of innovation?

Any infrastructure solution requires constant maintenance, monitoring, upgrading, and patching. By moving to the cloud, many of these tasks no longer need to be the concern of your team. Thus, you can free up those resources to work on developing specialized pipelines, doing cutting-edge analyses, and bringing together interesting diverse data sets—without being hampered by system downtime, unplanned maintenance, or delays getting the latest infrastructure versions in place.

Finding answers in the scale, efficiency, and security of the cloud

Researchers must ask these questions when evaluating their own DIY solutions to ensure not only effectiveness, efficiency, and scientific expertise, but also to understand how they're keeping their organizations safe from security and compliance risks. Many institutions are unaware of the enormous burden this becomes when they begin building homegrown solutions.

A flexible, easily scalable cloud-based environment for working with disparate and complex data sets, complete with out-of-the-box and custom workflow tools for advanced analysis, is becoming a hard-set necessity. Researchers simply can neither afford to use infrastructure that's quickly becoming obsolete nor waste time keeping up with all the nuanced security and compliance regulations.

"While cloud computing was not invented with science and genomics in mind—its major users are technology companies and other businesses—there is a growing list of cases where cloud computing helped to achieve important scientific goals."

BEN LANGMEAD AND ABHINAV NELLORE
"Cloud Computing as a Platform for Genomic Data Analysis and Collaboration,"
National Institutes of Health

Is DIY the best use of your resources?

If you're focused on advancing precision medicine as quickly as possible, why insist on spending your time and scientific expertise on building, maintaining, and updating infrastructure to fit your needs, while also keeping abreast of complicated security and compliance regulations? Especially if a solution exists that eliminates the heavy lift and security risk.

DNAnexus + Microsoft Azure: Putting the genomics cloud to work

You can achieve the global scale, efficiency, and competitive advantages of big data—without the tedious lift of DIY or the risk of security violations—by leveraging the flexible, end-to-end biomedical informatics solution from DNAnexus. The cloud-based platform enables the management of disparate data, such as genomic, multi-omic, and phenotypic in a secure, collaborative environment—at a global scale.

DNAnexus provides researchers the ability to push toward scientific breakthroughs, eliminating the need to spend resources on planning, building, and maintaining infrastructure.

DNAnexus leverages Microsoft Azure cloud, the industry's most compliant cloud environment spanning the industry's largest global infrastructure with 54 global Azure regions to date.

The partnership between DNAnexus and Microsoft provides researchers with a secure, scalable, and collaborative platform on top of a flexible and enterprise-ready cloud environment—propelling the application of genomics within precision healthcare.

The advantages of cloud-based next-generation sequencing analysis

Frost & Sullivan named DNAnexus one of the most significant enabling technologies helping usher in the era of large-scale genomics projects that have expanded to millions of genomes. DNAnexus leverages Microsoft Azure to deliver a powerful, purpose-built platform for enterprises that require scalable and secure data science solutions.

SCALABILITY AND REPRODUCIBILITY

Researchers don't have time to waste on the complexities of managing IT infrastructure and need to plan for future growth with confidence. DNAnexus Titan™ removes the heavy lift associated with scaling cloud-based NGS analysis by solving infrastructure challenges and increasing efficiencies. The solution provides not only preconfigured pipelines with preset parameter controls, ensuring the consistency of sample analysis, but also the flexibility to import your own tools into the platform. Plus, all data, tools, and workflows are tracked and version controlled for auditability and reproducibility, streamlining the analysis process.

> "The cloud's elasticity allows users to scale computing resources in proportion to the amount of data being analyzed, sidestepping constraints imposed by local clusters."
>
> BEN LANGMEAD AND ABHINAV NELLORE
> "Cloud Computing as a Platform for Genomic Data Analysis and Collaboration,"
> National Institutes of Health

FLEXIBILITY AND TECHNOLOGY INTEGRATIONS

Offering out-of-the-box features plus the flexibility to leverage custom workflows in tandem, DNAnexus Titan allows researchers to port, optimize, and deploy their own tools and pipelines with ease. The solution also allows for the optimization of customer workflows with industry-standard tools available on the platform.

Genomic analysis workflows can be easily developed, automated, and optimized to meet the growing global demand for NGS data while also minimizing the risk associated with security, privacy, and regulatory controls that apply to clinical and drug discovery pipelines.

The platform empowers team members to share data and tools in a secure environment in compliance with ISO 27001, HIPAA, CAP/CLIA, GxP, and GDPR, and was purpose-built for seamless collaboration. It also offers open API and SDK capabilities that enable pipeline portability, as well as a development cloud-based workstation, which lets collaborators work closely with their data and refine their applications.

SMOOTH TRANSITION WITH HELP FROM DNANEXUS EXPERTS

DNAnexus' in-house bioinformatics, cloud, and security experts help you easily connect to your existing technology stacks and accelerate your time to market with validated pipelines. The DNAnexus xVantage Group™ works with you to understand the dynamics of your company and get you up and running on the DNAnexus platform, providing seamless expert support for importing, deploying, and optimizing your data and workflows.

Cost match-up: DNAnexus vs. DIY

It's a common belief that buying a cloud-based genome informatics solution, such as DNAnexus, is more expensive than building one. However, this view doesn't take into account the total cost of execution and ownership.

When factoring in maintenance costs, the total cost of ownership for DNAnexus can turn out to be much lower than an internally built system, ranging from 20% to as much as 80% in savings. This is especially true in the long term, particularly if you consider that with DNAnexus, internal resources can be focused on other priorities than maintaining a homegrown solution.

Microsoft in healthcare

Healthcare organizations worldwide have long trusted Microsoft to provide the IT services they need to run their operations. From payers and providers to pharma and life-science researchers and other healthcare entities, enterprise-grade Microsoft systems have helped people work efficiently, maximize resources, and optimize patient care.

As a strategic solutions provider to global enterprises and government agencies, Microsoft has held a strong position on security and regulatory compliance. That history, position, and expertise is manifested and continually expanded in Microsoft Azure cloud services.

For the healthcare industry, Microsoft proves its commitment to compliance through:

- Microsoft Azure Security and Compliance Blueprint—HIPAA Health Data and AI
- Office 365 HITRUST CSF certification
- Minimum Acceptable Risk Standards for Exchanges (MARS-E) 2.0 Framework

As you move from your current Microsoft foundation to the cloud, Microsoft is a leading choice for implementing a configuration that best suits your operations wherever you are in your cloud journey: hybrid cloud, public/private cloud, or 100% cloud infrastructure.

DNAnexus | Microsoft Azure

Less maintenance, more breakthroughs

Breakthroughs are hard enough to make without having to maintain unnecessarily complex proprietary solutions. Hours spent on the upkeep of DIY solutions are better spent focusing on what's truly important—scientific breakthroughs that will have a meaningful impact. The DNAnexus platform will help your organization future-proof your investments and scale with the velocity of technological innovation.

Together with Microsoft Azure, DNAnexus provides the world's most secure and trusted cloud-based NGS informatics platform consistently at the cutting-edge of scientific breakthroughs. Join us in accelerating scientific collaboration and discovering how to tackle the world's most exciting opportunities in human health…and beyond.

EXPLORE MORE

[1] Tufts Center for the Study of Drug Development, **"Personalized Medicine Is Gaining Traction, but Faces Multiple Challenges,"** press release, May 14, 2015.

[2] Zachary D. Stephens, Skylar Y. Lee, Faraz Faghri, Roy H. Campbell, Chengziang Zhai, Miles J. Efron, Ravishankar Iyer, Michael C. Schatz, Saurabh Sinha, Gene E. Robinson. (2015, July 7). **"Big Data: Astronomical or Genomical?"** Retrieved from *PLOS Biology.*