kainos®

GENOMICS IN THE CLOUD

Driving efficiencies to advance scientific progress

Contents

The growth of genomics sequencing

The last 20 years have seen tremendous growth in human genomics. To put things in perspective, the Human Genome project began in 1990 as an international scientific research project with the goal of creating the first human genome sequence.

The project took 10 years to create its first working genome sequence draft and 13 years before it was completed. Today, this process can be completed in under 24 hours[1].

Due to these incredible advances in technology, millions of genomes have been sequenced so far, making it easier to study diseases associated with mutations in a single gene.

But with genomics sequencing analysis becoming more accessible, the amount of data being produced is rapidly expanding, with the amount of raw genomic data being produced around the world doubling every seven months[2].

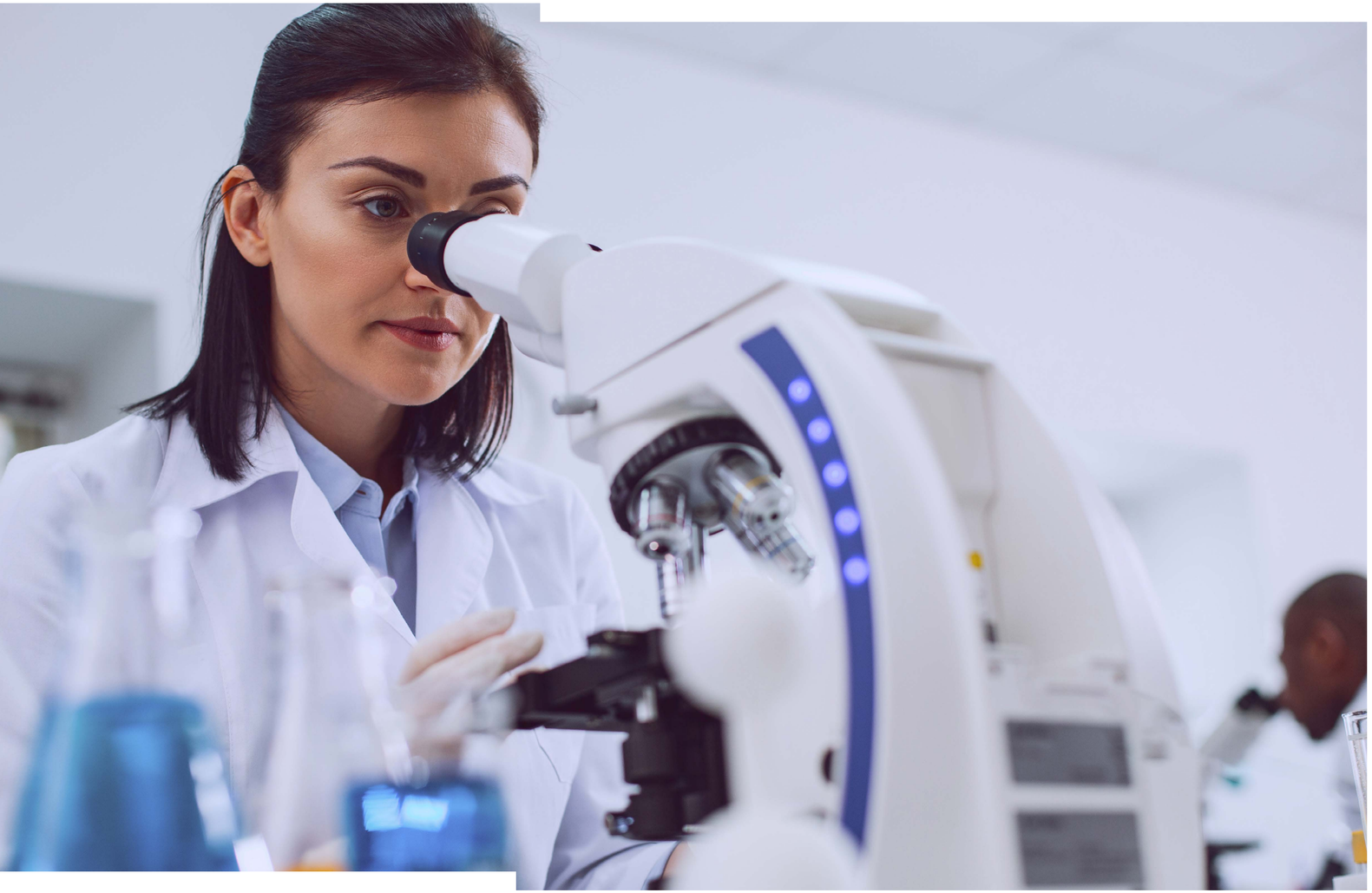
By 2025 it is estimated that:

Between 100M and 2B genomes will be sequenced[3]

Between 2 and 40 exabytes of storage capacity will be required to store the entire globe's human genomic data[4]

This data, combined with ever-growing amounts of single-cell and functional genomics data, digital medical records, and other critical biomedical data, has the potential to substantially enhance our understanding of the fundamental processes for healthy life and revolutionise the treatment of disease. But this doesn't come without its challenges.

Data challenges – overcoming research bottlenecks



Genomics data output is increasing all the time. But while these ever-larger scientific datasets may be a goldmine for discovery, analysing them within on-premise legacy environments – which many life-science organisations still use – has become a bottleneck in genomics research. Massive processing power and scalability is required, presenting a challenge to organisations of all sizes with on-premise storage systems based on outdated legacy infrastructure.

For example, to gain insights from huge and archived datasets, a researcher must secure sufficient storage space and perform large, time-consuming downloads, followed by a compute-intensive data re-analysis from scratch.

Many labs aren't equipped for this, so valuable data goes unused. In addition, the velocity and volume of genomic data continues to rise in response to reduced sequencing costs and broader adoption. Eventually, single organisations may struggle to independently manage, sequence, process and analyse all insight available from a particular data set.

Instead, we may see smaller, agile groups capable of looking at specific problems that drive the development of insight. They will however require the ability to easily and securely access this information – which is only achievable with modern cloud-based technology.

DATA CHALLENGES

Let's look at some of the other key areas of consideration when it comes to the most common data challenges.

Data challenges – overcoming research bottlenecks

Key areas of consideration

The need for greater data sharing

The National Human Genome Research Institute (NHGRI) strongly encourages studies involving human data to use data generated from sources with participant consent, for unrestricted access, or for general research use with controlled access.

At the same, time, life-science organisations seek to explore genomic science and its impact on research, health, and society, further accelerated by an increasing number of government-funded genome projects. The European Molecular Biology Laboratory (EMBL-EBI) for example offers data that is used extensively across the world by more than five million researchers in academia and industry, with some 64 million data requests made daily to its websites[5].

This demands close collaboration across the global scientific community along with fast, secure open access to biological research. However, uploading data into on-premise repositories can be time consuming and may result in data that's minimal and sparse, should researchers only deposit what's required to remain compliant. Sometimes data may be stored in more than one place, which again creates challenges associated with time spent locating and accessing information.

Advancing technology

Wearable devices powered by the Internet of Things (IoT) have the potential to greatly impact biomedical research, by providing scientists around the world access to data that supports the advancement of precision medicine.

For example, in August 2020, Quest Diagnostics launched an automated next generation sequencing (NGS) engine to power Ancestry Health. This now enables people to access precise genetic testing and gain insights into inherited diseases, including cancers of the colon and other conditions[6].

However, each new device or app generates gigabytes, potentially even terabytes of data every day. This data needs its own back-end capability for sending, requesting, and processing information on a massive scale, which is stretching the limits of hardware, software, and datacentres.

Genomics in the cloud

When it comes to effectively sharing and analysing large-scale data sets, cloud computing plays a central role. Genomics healthcare is reliant on leveraging large sequencing data sets, requiring the use of large-scale computational resources. That's why cloud computing is ideally suited for these workloads given the scalability, repeatability, and granular control of security it provides.

With cloud technology, life-science organisations can enable the management and storage of such data sets with the ability to provide on-demand technology – minus the need to invest heavily in physical hardware. As well as storage, cloud technology greatly enables the democratisation required to access meaningful insights.

Data access

Compute power

Analytical tools

Private and secure digital workspaces

Cloud technology makes it possible for collaborators to make optimal use of their own customised workflows, applications, and datasets. Researchers can access their workspace from anywhere in the world and be supported to work with greater flexibility and agility in the new hybrid world of work.

A facilitator for collaboration

When it comes to collaboration, cloud technology can democratise access to the storage and compute-intensive tools required to fully leverage genomic data. This makes sharing large amounts of data easier to achieve, with the distributed nature of the cloud making it a natural venue for driving and facilitating collaborative effort.

Genomics in the cloud

Greater computational power and analysis

The inherent scalability and flexibility of the cloud also lends well to genomic data processing, storage, and analysis. Resources can be adjusted on-demand to meet the requirements of an organisation, individual or workload.

Stronger data protection

Given the highly sensitive nature of biometric and medical data being handled and the criticality of its use, security and stability is crucial. With the right cloud-based tools and services, researchers can securely store, process, explore, and share large genomic datasets with end-user privacy safeguards and secure communications between services. From data transfer and storage to aggregation and governance, cloud technology can leverage state-of-the-art security features to protect sensitive data and help ensure organisations stay compliant regardless of where data resides.

Storage and performance gains

Cloud providers maintain data centres in a way that achieves economies of scale and removes the worry associated with outages, software patches, service contracts, and damaged parts. This minimises downtime, which helps maintain continuity when it comes to essential work and research.

Lower costs and maintenance

With no need to invest in or maintain computer storage hardware, organisations can reduce the need for capital investments in hardware and related operational costs – which allows them to re-focus valuable time, money, and resource where it's needed most.

Reproducibility and delivering at speed

Cloud lends itself well to automation, which makes it possible to quickly deploy identical environments at the touch of a button. This removes management overheads and delays, and instead allows researchers to focus on what's important, research.

Key cloud challenges and mitigations

The benefits of moving to the cloud are many, and with it comes several key challenges that prevent life-science organisations from taking the first step.

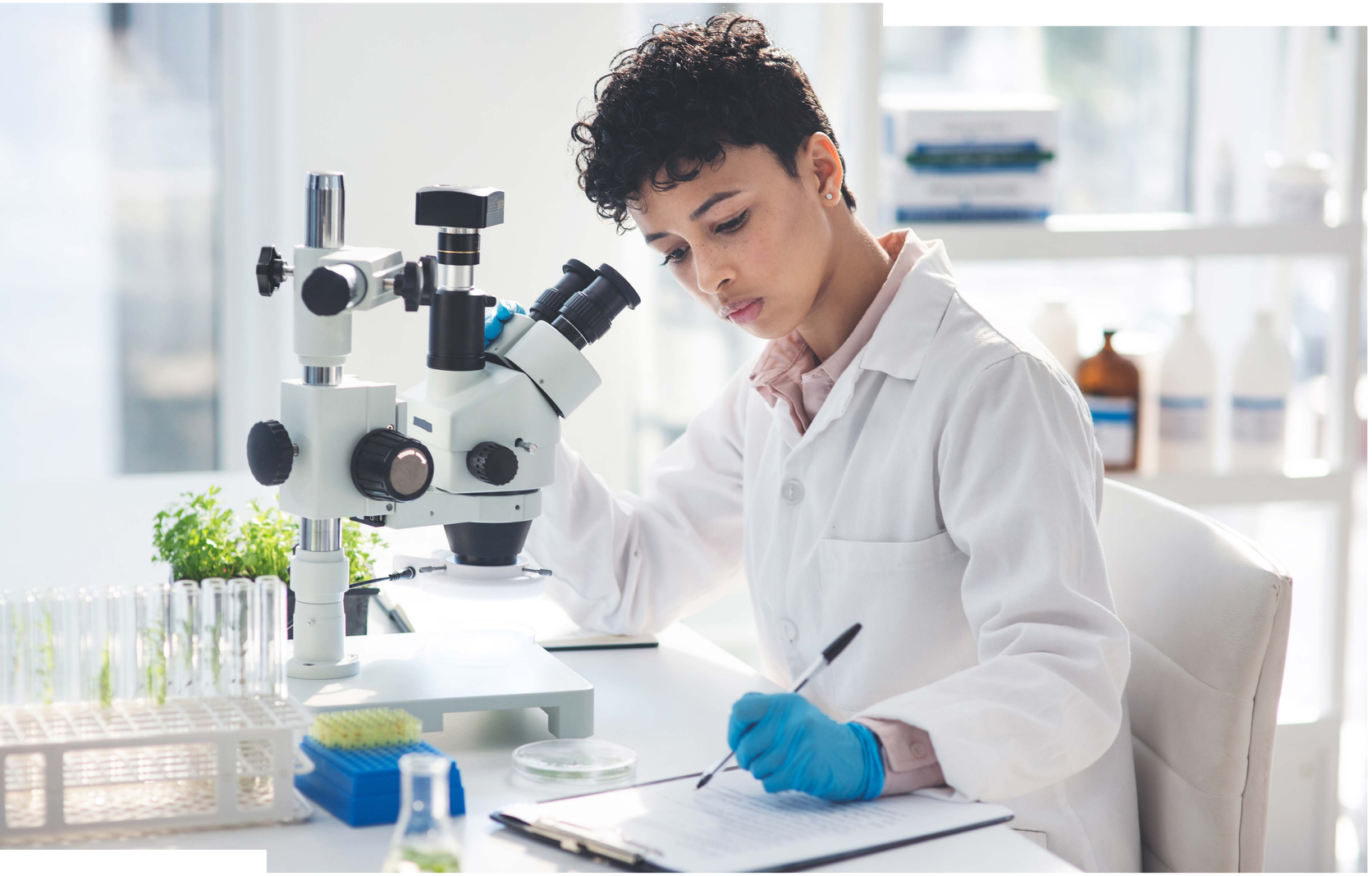
Cloud vs On-Premise costs - although building and maintaining an on-premise solution may seem like a financially viable option, since initial costs can be amortised over its lifetime, there is a far greater cost vs power ratio in cloud technology. Funding is also available to help reduce any initial migration and adoption costs.

Data privacy concerns - due to privacy standards such as those set out by The Health Insurance Portability and Accountability Act (HIPAA) and The database of Genotypes and Phenotypes (dbGaP), many organisations have already built compliant systems in the cloud. The challenge they face is convincing other bodies, such as Internal Review Boards or IT administrators, that standards are being followed. Demonstrating ISO compliance certifications, developing a trust and governance framework, and undertaking independent verification of security, privacy, and compliance controls are some of the steps that can be taken to change perspectives, and ensure regulatory compliance is being adhered to.

Poor data quality - not all datasets, whether raw data or metadata, are of high quality or annotated with informative (or even correct) metadata. In the absence of reliable and automatic methods for dealing with poor quality data and metadata, researchers must approach public data cautiously with the right processes and stringent guardrails in place.

Cloud skills gaps - the task of recruiting and retaining people with specialist cloud skills can be a costly and time-consuming task. However once trained, engineers can work faster and smarter – especially when equipped with the right automation tools. From an expense perspective, funding is often available to help with the costs of cloud training and certifications.

Kainos - the No 1 digital provider for healthcare migrations



There's no doubt that cloud technology can help solve many challenges for life-science organisations who deal with huge amounts of sensitive health data. At Kainos we not only recognise the value in moving to the cloud, we understand the difficulties some organisations face in managing poor datasets, a lack of cloud skills, and regulatory compliance requirements.

Our focus is on supporting our customers to make this critical transformation, by combining a design-led approach with a highly capable skill set, **which has enabled us to become one of the NHS key cloud migration partners.**

BRINGING OUR EXTENSIVE KNOWLEDGE TO THE GENOMICS SPACE

With experience building some of the UK Government's most sensitive public applications using our true application and cloud engineering capabilities, we migrate complex legacy platforms, and have been successful in unlocking the power of big data for Genomics England (GE) and many other healthcare clients.

Our wealth of expertise across artificial intelligence (AI), intelligent automation (IA), data, design, and digital transformation are woven into our cloud solution.

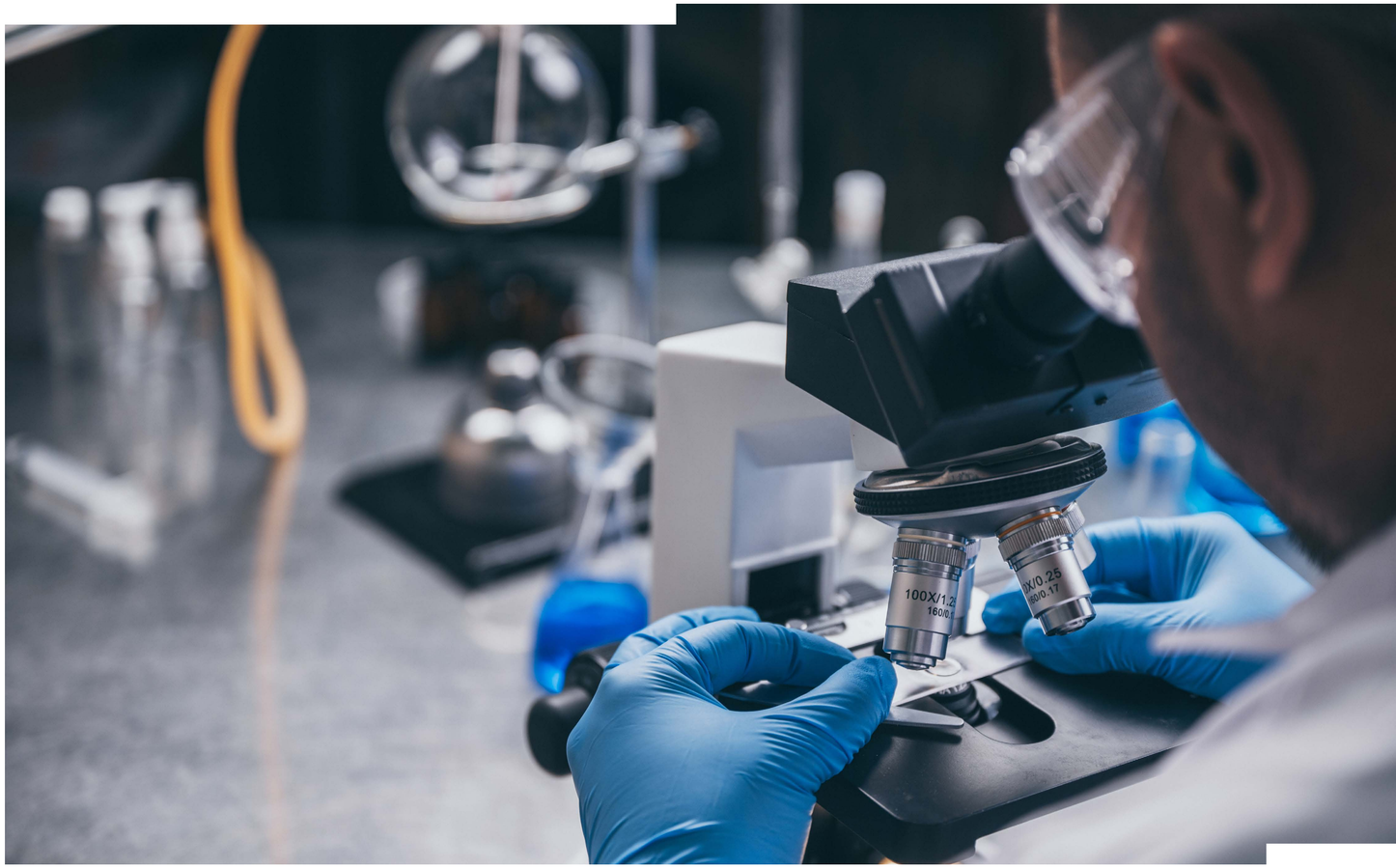
Bringing genomics healthcare to life for GEL

Genomics England (GEL) is one such client we helped bring to life genomic research applications, for the benefit of human health. Founded to run the 100,000 Genomes Project in the UK, GEL is now driving forward genomic research applications for human health.

As a trusted Amazon Web Services (AWS) Professional Services partner we were tasked to help plan, deploy, and deliver an environment that now provides approved researchers with the tools, storage, databases, and platforms needed to enable scientific discovery and accelerate patient care.

Working in close partnership with GEL and AWS, within 6 months we achieved:

- **99% reduction in time taken to perform common tasks**
- **Greater security and durability of hundreds of thousands of genomic data samples**
- **A migrated environment estimated to cost a tenth of previous cost**
- **Optimised data management platforms for unique needs of genomic datasets**
- **Democratised genomic research access**



> As we continue to progress our work and generate more robust data sets, access to elastic storage and compute services will enable our organization and the research community to access and analyse data securely and cost-effectively. By hosting on AWS, we can democratise access to our data.
>
> **Pete Sinden, Chief Information Officer, Genomics England**

Why partner with Kainos

Whether you need to transfer large volumes of genomic data to the cloud, reduce long term data storage costs, or collaborate securely across the globe, we're here to help.

Here's how:

Cut costs and drive performance

Our default approach to cloud migration has cloud cost optimisation at its core. Allowing you to continuously optimise spend while building modern, scalable and performant applications. Moving to the cloud should help lower the total cost of ownership (TCO) through reduced hardware, cost-effective software acquisition, efficient compute utilisation, and overall reduced datacenter management and footprint.

Better aggregated and governed data

We can deliver solutions that simplify the process of aggregating, storing, and analysing your genomics datasets, while helping you better govern your information and quickly link it with other types of medical data. As you welcome more collaborators, we enable you to maintain data integrity with data access controls and permissions suited to your needs.

Automation that powers innovation

Boost workflows and reduce time spent on analysis with our powerful workflow automation solutions. We provide the foundation for powerful machine learning and high-performance compute resources, capable of transforming genomic data into biological insight.

Drive efficiency at-scale

By moving your data sets to a compliance-ready cloud environment we can help you find greater computing efficiency at-scale, achieve reproducible data processing, and develop data integration capabilities that accelerate insight and new correlation discovery.

Make faster clinical decisions

With your data in the cloud you will be able to access insights from your patient's genetic codes, your clinicians can make more accurate decisions when it comes to diagnoses and ultimately – better, faster treatment decisions. Industry leaders around the world rely on AWS and AWS Partners secure and compliant solutions to implement genomics in a clinical setting.

Secure global collaboration

We offer cloud migration services that are reliable and consistent and move you onto platforms that provide high data availability while adhering to industry regulations. Our security standards and compliance certifications include HITRUST, GDPR compliance, FedRAMP, HIPAA, ISO 27001, ISO 9001, and ISO 3425.

What's next for cloud genomics

While cloud computing has changed how life-science organisations manage their data and resources, at the same time it is changing how researchers collaborate and work with vast and critical datasets. As the cloud makes increasingly more important inroads into genomics, it is important for researchers to be able to fully harness the new modes of analysis and collaboration it enables.

Kainos has the tools and solutions to help you achieve this, by migrating and securely storing your genomic data in the cloud. Partner with us to accelerate analysis and enable global collaboration, to ensure your impact and reach is as wide as possible.

IF YOU'RE READY TO ADVANCE YOUR CLOUD JOURNEY, BOOK A 1-1 MEETING WITH A CLOUD MIGRATION EXPERT TO SEE HOW WE CAN HELP YOU CONTINUE DRIVING THE NEXT WAVE OF GENOMIC INNOVATION

[1] Genomics England
[2] Sanger Institute
[3] AWS
[4] AWS
[5] Wellcome Genome Campus
[5] Quest Diagnostics

kainos®

Find out more at **kainos.com**

Follow our story:
@KainosSoftware
KainosSoftware
search 'Kainos'