DATA SHEET

Illumina Connected Analytics

Transform your bioinformatics operations

- Streamlined integration with sequencing system
- Scalable throughput to meet sample processing needs
- Intuitive interface to select, build, and customize workflows

illumina®

Introduction

Advances in next-generation sequencing (NGS) technologies have dramatically changed the rate at which life sciences and clinical research is conducted. As the speed of sequencing increases and the cost decreases, the ability to generate data will far outpace the ability to extract biological and clinical insight from the data. Addressing the challenges of pipeline development and implementation, scaling informatics workflows, and maintaining secure data management requires a flexible and comprehensive platform. Illumina Connected Analytics allows users to build, version, and deploy flexible analytical pipelines while maintaining data privacy, security, and compliance.

Illumina Connected Analytics is a secure, cloud-based bioinformatics data platform for operationalizing informatics and driving scientific insights (Figure 1, Table 1). Illumina Connected Analytics enables customers to create workflows that are:

- Connected, featuring streamlined integration with Illumina systems and software
- Scalable, with the ability to manage, analyze, and query massive data sets
- Flexible, allowing users to build and customize workflows with DRAGEN™ applications and custom analysis pipelines

Streamlined workflow

Illumina Connected Analytics is a central component for labs performing NGS studies with Illumina sequencing systems. Taking advantage of the elasticity of resources afforded by cloud computing, Illumina Connected Analytics supports operations at any scale, from occasional screening to tens of thousands of cells in complex single-cell projects to population-scale whole-genome sequencing, with the same architecture. Users can seamlessly integrate their instruments with Illumina Connected Analytics.

Within Illumina Connected Analytics, data can be automatically analyzed with ready-to-use DRAGEN secondary analysis pipelines or custom pipelines, depending on the specified workflow. The broad range of analysis options spans quality control to data aggregation and advanced data science tools for rapid, scalable data processing. Illumina Connected Analytics provides an extensible platform with a rich set of RESTful application program interfaces (APIs) and a command-line interface (CLI) tool. These APIs maximize the efficiency of workflows as data are transferred, accessed, and used across its lifecycle.

Table 1: Illumina Connected Analytics at a glance

	Feature	Benefit
Security and privacy	Compliance	Developed under the Illumina Quality Management System (QMS) with adherence to local, regional, and global regulatory standards including GDPR,[2] HIPAA,[3] ISO 27001/27701,[4,5] and ISO 13485[6]
	Security controls	Maintain strict data segregation, "in-transit" (TLS 1.2) and "at rest" (AES 256) encryption
	Audit trail	Maintain activity and event log tracking for access monitoring
	Multifactor authentication (MFA)	Apply institutional credentials to manage and control system access
Resourcing	Scale-on-demand	Scale cloud storage and compute needs to meet the current level of demand
	Compute resources on demand	Reduce costs by paying only for compute resources in the pipeline engine
Management	Project and user management	Manage user access and activity for granular privacy
	Data sharing	Bridge data silos to foster global collaboration and support data delivery
	Data archive	Reduce costs for long-term data management with archival storage tiers
	Bring your own AWS cloud bucket	Access data stored within a privately managed AWS cloud account
Usability and accessibility	Sequencing system integration	Stream data seamlessly from Illumina sequencing systems to Illumina Connected Analytics storage and analysis
	Tools and pipelines	Apply out-of-the-box DRAGEN pipelines or import existing tools
	APIs and CLI	Interact with the platform, tools, and data programmatically via CLI or RESTful API
	Intuitive graphical interface	Access data and tools through a graphical web interface; display data with R and Python packages
Advanced tools	Docker, Nextflow, and CWL support	Write pipelines in common workflow language and launch analyses in the cloud with ease
	Integrated with JupyterLab	Run custom scripts, build and train AI/machine-learning models, and interact with data within the platform through collaborative notebook workspaces
	Data aggregation and query	Organize and query structured multiomic data sets to power tertiary analysis

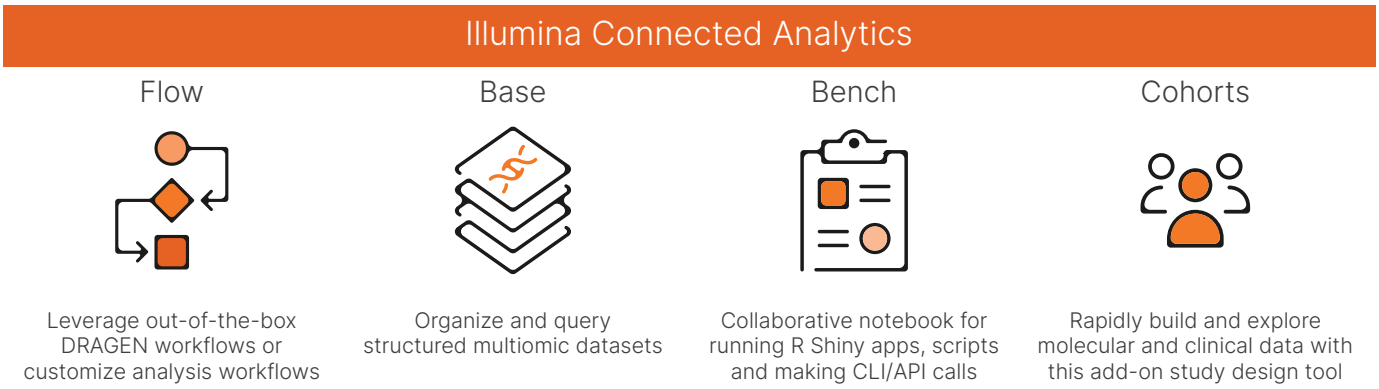

Figure 1: Illumina Connected Analytics modular components—Process, aggregate, and analyze data within a secure, access controlled, data management environment.

BaseSpace™ Sequence Hub

Included with all Illumina Connected Analytics subscriptions, BaseSpace Sequence Hub is a direct extension of your Illumina instruments. Increase lab efficiency with BaseSpace Sequence Hub integration by setting up runs, monitoring run quality, and streaming sequencing data directly to the cloud. Encrypted data flow from the instrument into BaseSpace Sequence Hub enables easy data management and analysis using a curated set of apps within a secure environment.

Transform reads to data

Illumina Connected Analytics offers various options for secondary data analysis, streamlining the reads-to-results workflow. With the flexibility to use ready-made pipelines or construct and configure customized pipelines, Illumina Connected Analytics can support virtually any informatics application.

Ready-to-use options

Illumina Connected Analytics delivers powerful out-of-the-box tools and pipelines for processing data, including access to DRAGEN secondary analysis,[1] which provides accurate, efficient, and comprehensive secondary analysis of sequencing data (Figure 2).

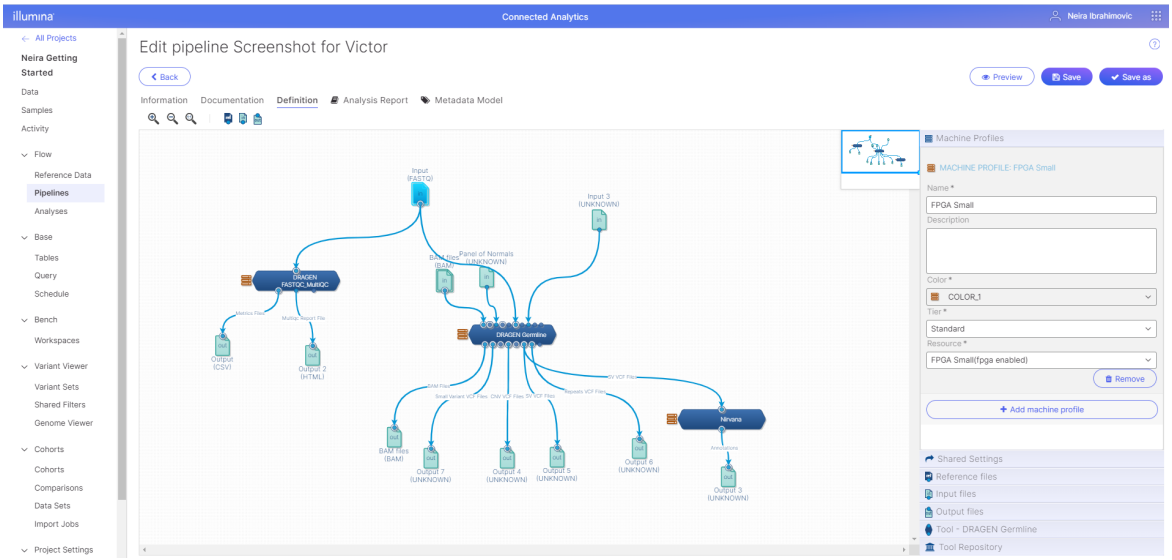

Figure 2: Illumina Connected Analytics Flow—Flexible building, deployment, and versioning of analytical pipelines.

Customizing pipelines

Bioinformaticians can import existing analysis tools from a Docker Hub image repository,[7] or construct and edit new pipelines using Nextflow, common workflow language (CWL), and the graphical pipeline editor. Lab operators and other scientists can launch pipelines with ease using the intuitively designed user interface.

Data management and control

With the increase in data generation comes a greater need for infrastructure to support sharing, reusing, and integrating data within the scientific community to amplify the value of individual data sets. To address this need, Illumina Connected Analytics incorporates several features designed to enable adoption of best practices in data management.

Access control

Fine-grained access control enables an administrator to set permissions and take advantage of existing institutional credentials to control access. An audit log serves as a record of events and changes, logging each user when they access the platform and their actions while using the platform, enabling enforcement of compliance and accountability.

Open format

Illumina Connected Analytics is designed as a data-agnostic platform. It supports analysis of multiple data types, including molecular, clinical, phenotypic, and unstructured data such as images.

Collaboration

Illumina Connected Analytics empowers collaboration across geographic boundaries in a compliance-preserving manner. Data and tools can be instantly delivered and shared with other users in a manner that preserves data integrity and privacy. In addition, data and analytical tools hosted in an external cloud source can be imported into Illumina Connected Analytics for analysis and sharing.

Data aggregation and querying

Illumina Connected Analytics automates complex aggregation and integration steps to create a functional knowledge management system that encompasses data from millions of samples. It captures virtually any type of data, genotypic, phenotypic, metadata, annotations, and other associated information, available. Users can define their own data models, write their own queries, and explore connections between the data sets as they need. Data aggregated on Illumina Connected Analytics represents a wealth of information that can be used to discover novel biomarkers, stratify patient populations, monitor assay performance over time, and more.

Use Cohorts to explore clinicogenomic data at scale

Illumina Connected Analytics Cohorts is a study design tool for rapidly building and exploring cohorts (Figure 3). Illumina Connected Analytics Cohorts accelerates the path from genomic discovery to meaningful studies and enables better decision making on study design by aggregating multiomic and phenotypic data in a single place.

Augment cohorts with included harmonized public data sets, including The Cancer Genome Atlas (TCGA), Broad Rare Genomes Project, 1000 Genomes Project, and Gene Expression Omnibus. Build hypotheses in minutes rather than weeks by leveraging an integrated cohort browser. Access highly curated public data sets. Use simple visual tools to select and visualize cohorts, avoiding the need for complex querying.

Secure notebook environment to drive insights

Algorithm development and customization are essential components of Illumina Connected Analytics for supporting deep data exploration. An interactive programming module, leveraging popular Jupyter notebooks, Python, and R empowers data scientists to analyze aggregated data in a seamless and secure environment (Figure 4).

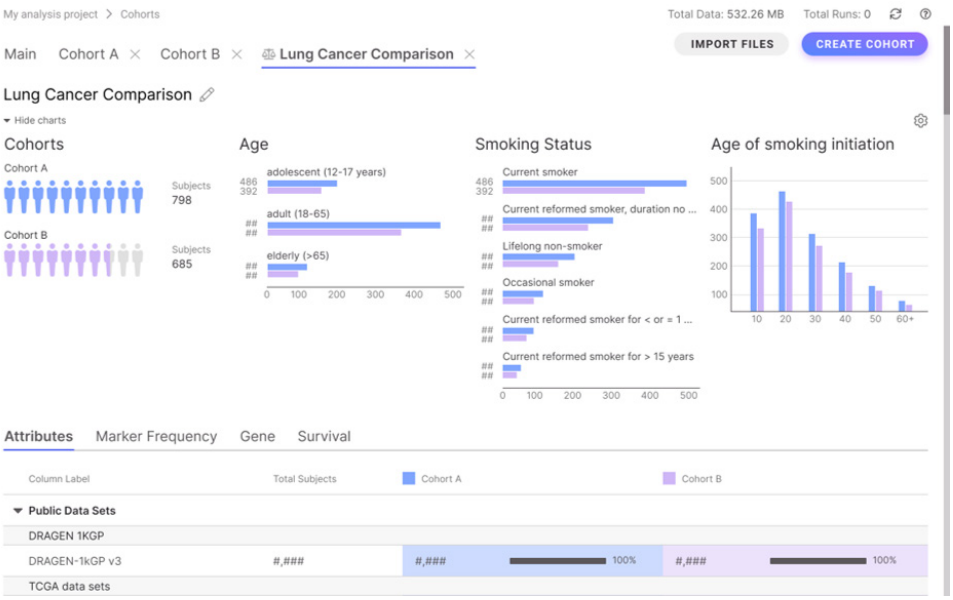

Figure 3: Illumina Connected Analytics Cohorts—Rapidly build and explore molecular and clinical data using this add-on module in Illumina Connected Analytics.

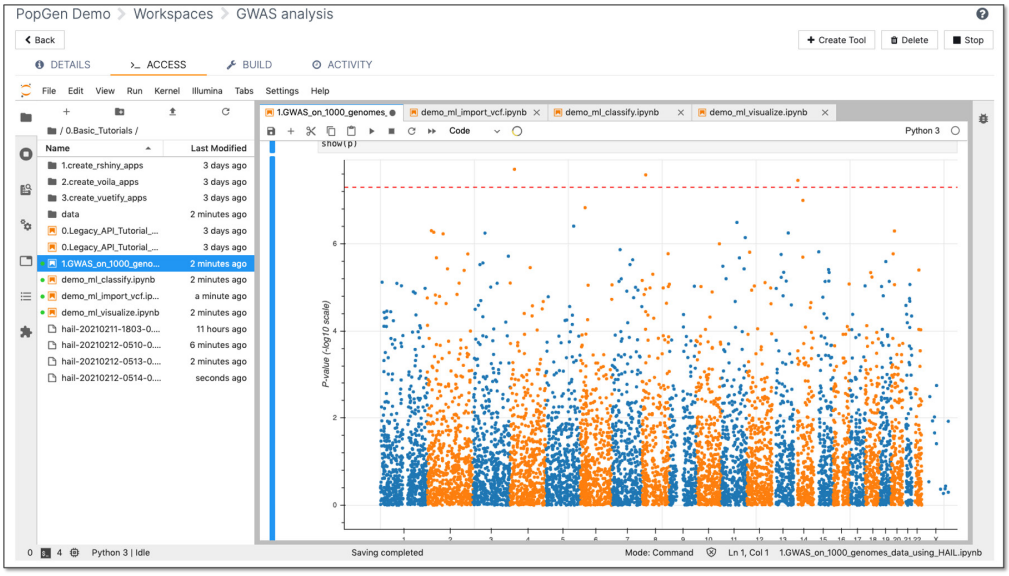

Figure 4: Illumina Connected Analytics Bench—Flexible and interactive computing tools are built into Illumina Connected Analytics Bench, including Jupyter notebooks, Python, and R.

In the method and algorithm development phase, users can develop or modify pipelines in a sandbox environment. There, users have access to standard libraries and can easily bring in their own custom libraries, such as TensorFlow[8] or scikit-learn,[9] to build custom and complex scripts to process data within Illumina Connected Analytics in a collaborative fashion. When users are ready to move to the production phase, Illumina Connected Analytics enables conversion of notebooks into tools. These tools will then be available in the Illumina Connected Analytics tools repository and incorporated into production pipelines.

Security and compliance at the core

Security is of paramount importance when operating with genomics data for research, clinical therapeutics, and human diagnostics. Illumina Connected Analytics employs various digital and administrative measures to meet even the most demanding data security requirements:

- Data uploaded from sequencing instruments are encrypted using the AES256 standard[10] and protected by transfer layer security (TLS)
- Data within Illumina Connected Analytics are hosted on Amazon Web Services (AWS) to maintain compliance with a wide variety of industry-accepted security standards by using AWS Well-Architected best practice[11]
- Illumina Connected Analytics is hosted in more than 10 global AWS regions to enable compliance with local regulations to maintain genomic data within particular geographic territories
- The authentication service is supported by security assertion markup language (SAML) 2.0 to manage institutional users and passwords (optional)
- Audit reports support traceability of data provenance

Illumina Connected Analytics also supports customers operating in regulated environments who must comply with stringent requirements. Illumina Connected Analytics was developed in accordance with the Illumina software development lifecycle process under the Illumina Quality Management System (QMS). Also, processes within the Illumina QMS have adopted industry best practices and relevant standards, including:

- International Organization for Standardization (ISO) ISO 27001:2013 information security management system[4] and ISO 27701:2019 security techniques[5]
- ISO 13485 organizational Quality Management System (QMS) best practices[6]
- General Data Protection Regulation (GDPR)[2]
- Health Insurance Portability and Accountability Act (HIPAA)[3]
- Guaranteed data residency to address local regulatory and compliance requirements

Ordering information

Product	Catalog no.
ICA Enterprise Annual Subscription	20038994
ICA Professional Annual Subscription	20044876
ICA Training and Onboarding	20049422
ICA Cohorts Annual Subscription	20065842
ICA Cohorts Annual Subscription	20065842
Illumina Analytics - 1 iCredit	20042038
ICA Enterprise Srvc and Compliance Add-on	20066830

Abbreviations: ICA, Illumina Connected Analytics

Learn more

Illumina Connected Analytics

References

1. Illumina. Illumina DRAGEN secondary analysis. www.illumina.com/products/by-type/informatics-products/dragen-secondary-analysis.html. Accessed March 1, 2024.
2. General Data Protection Regulation (GDPR) Compliance Guidelines. GDPR website. gdpr.eu. Accessed March 1, 2024.
3. US Department of Health and Human Services. Health Information Privacy. HHS website. hhs.gov/hipaa/index.html. Accessed March 1, 2024.
4. International Organization for Standardization. ISO-ISO/IEC 27001—Information security management. ISO website. iso.org/isoiec-27001-information-security.html. Accessed March 1, 2024.
5. International Organization for Standardization. ISO/IEC 27701:2019—Security techniques. iso.org/standard/71670.html. Accessed March 1, 2024.
6. International Organization for Standardization. ISO 13485—Medical devices. ISO website. iso.org/iso-13485-medical-devices.html. Accessed March 26, 2024.
7. Docker. Docker Hub Image Container Library. https://hub.docker.com. Accessed March 26, 2024
8. TensorFlow. TensorFlow website. tensorflow.org. Accessed March 1, 2024.
9. scikit-learn: machine learning in Python. scikit-learn website. scikit-learn.org/stable/. Accessed March 1, 2024.
10. National Institute of Standards and Technology. Advanced Encryption Standard (AES). nist.gov/publications/advanced-encryption-standard-aes-0. Accessed March 26, 2024.
11. Cloud Security—Amazon Web Services (AWS). Amazon website. aws.amazon.com/security. Accessed March 1, 2024.
12. Illumina. iCredits for Data Storage & Analysis. illumina.com/products/by-type/informatics-products/icredits.html. Accessed March 1, 2024.

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-00684 v3.0.