DATA SHEET

Illumina Connected Analytics

Production informatics workflows at scale

- Import, build, and edit workflows with tools like CWL (Common Workflow Language) and Nextflow
- Organize data in a secure workspace and share it globally in a compliant manner
- Interpret data in a flexible computing environment that includes JupyterLab Notebooks

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. Meeting the challenges of secure data management, scaling up infrastructure, and building and deploying new informatics workflows requires a flexible and comprehensive platform. Illumina Connected Analytics (ICA) allows users to build, version, and deploy flexible analytical pipelines while maintaining data privacy, security, and compliance at scale.

ICA is a secure genomics data platform to operationalize informatics and drive scientific insights (Figure 1, Table 1). ICA enables users to:

- Build and customize analysis pipelines
- Execute production workflows at scale
- Explore and share data and results

Streamlined workflow

ICA is a central component for labs performing NGS studies with Illumina sequencing systems. Taking advantage of the elasticity of resources afforded by cloud computing, ICA 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 ICA.

Within ICA, data can be automatically analyzed with ready-to-use DRAGEN™ 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. ICA 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, and include Global Alliance for Genomics and Health (GA4GH)-compliant APIs.[1]

Table 1: ICA at a glance

	Feature	Benefit
Security and privacy	Compliance	Adhere to local, regional, and global regulatory standards, HIPAA and GDPR standards, and ISO 27001 certifications
	Security controls	Maintain strict data segregation, "in-transit" (TLS 1.2) and "at rest" (AES 256) encryption
	Audit trail	Maintain an activity log tracking who accessed what data and when
	Single sign-on (SSO) (optional)	Leverage institutional credentials to control access
Resourcing	Compute resources on demand	Reduce costs by paying only for compute resources in the pipeline engine
	Scale on-demand	Scale cloud storage and compute needs to meet current level of demand
	Platform and usage dashboard	Display resource demands visually for understanding, managing, and anticipating needs efficiently
Management	Project and user management	Manage user access and activity for granular privacy
	Data sharing	Bridge data silos for large-scale, global collaboration
	Data archive	Reduce costs by archiving unused data in lower cost storage tiers
Usability	Sequencing system integration	Flow data seamlessly from Illumina sequencing systems
	Visual pipeline builder	Create pipelines without writing code
	Tools and pipelines	Leverage out-of-the-box pipelines and import custom tools
	APIs and CLI	Interact programmatically with the platform using tooling based on user preference
	"Bring your own bucket"	Access data stored within a privately managed cloud account
	Data visualization	Create dynamic visual plots and interactive web apps to 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
	RESTful, GA4GH-compliant APIs	Enable programmatic access to tools and data and interoperability with other software environments
	Integrated with JupyterLab	Perform advanced data analytics; build and train AI/ML models with R and Python
	Data aggregation and query	Perform population-level data queries using SQL

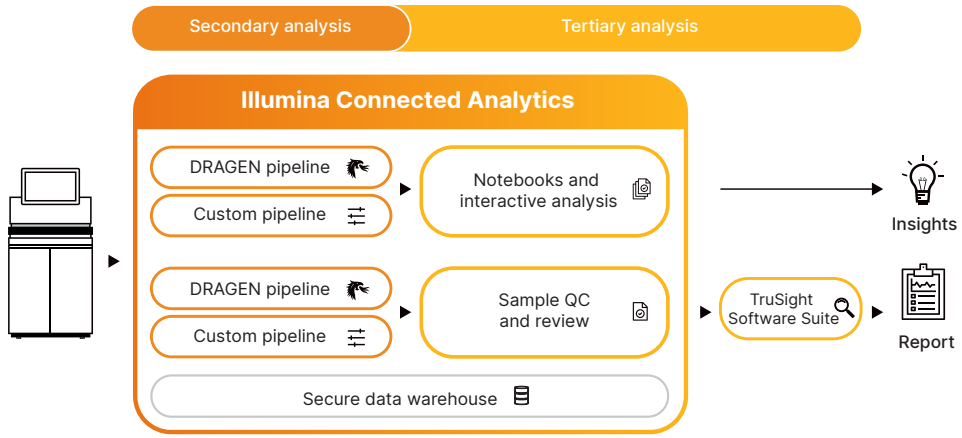

Figure 1: ICA forms the foundation for data management and analysis

Transform reads to data

ICA 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, ICA can support virtually any informatics application.

Ready-to-use options

ICA delivers powerful out-of-the-box tools and pipelines for processing data, including access to the DRAGEN Bio-IT Platform,[2] which provides fast, accurate secondary analysis of sequencing data (Figure 2).

Customizing pipelines

Bioinformaticians can import existing tools from a docker image repository, or construct and edit new pipelines using Nextflow, CWL, and the graphical pipeline editor. Lab operators and other scientists can launch pipelines with ease using the intuitively designed user interface.

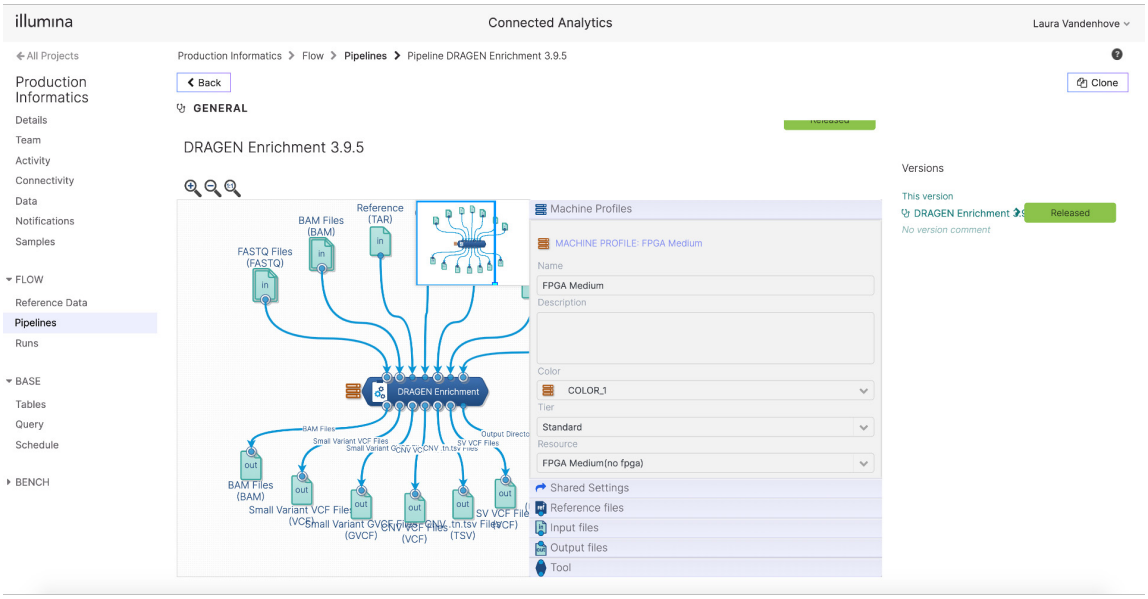

Figure 2: DRAGEN pipeline in ICA—Ready-to-use DRAGEN pipelines in ICA enable fast, accurate, reads-to-report secondary analysis.

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, ICA 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

ICA 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

ICA 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 ICA for analysis and sharing.

Data aggregation and querying

ICA automates complex aggregation and integration steps to create a functional knowledge management system that encompasses data from millions of samples (Figure 3). 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 ICA represents a wealth of information that can be used to discover novel biomarkers, stratify patient populations, monitor assay performance over time, and more.

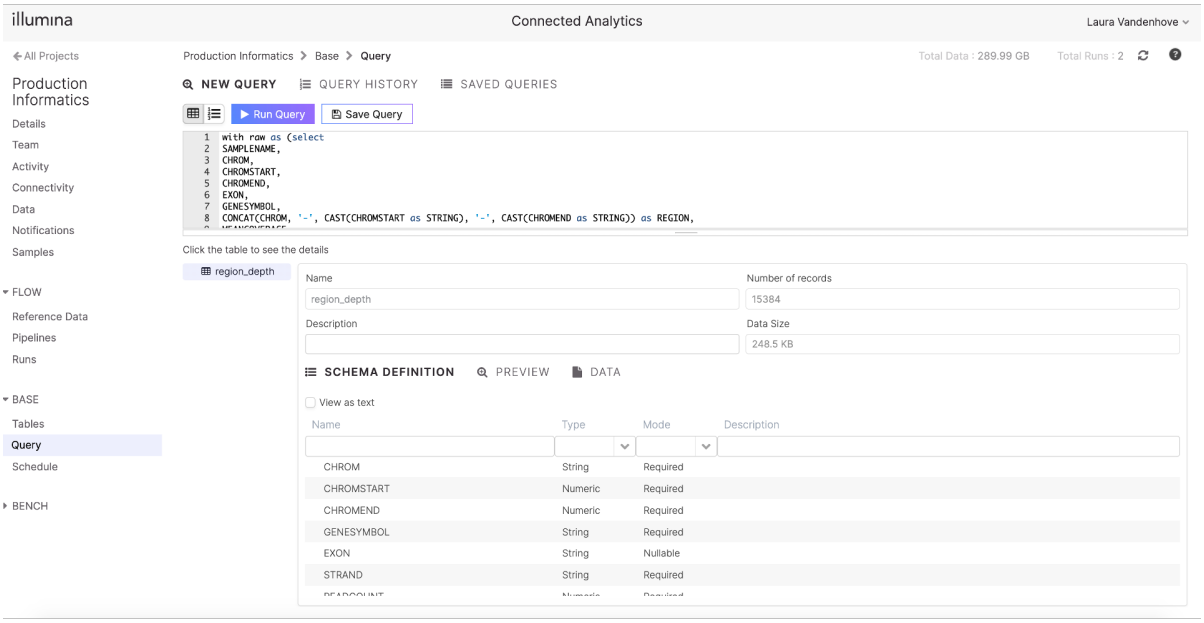

Figure 3: ICA enables data aggregation, mining, and continuous learning—Users can explore connections between data sets to answer user-driven questions.

Secure notebook environment to drive insights

With the myriad of ongoing data exploration, the ability to develop and customize algorithms is essential. An interactive programming module, leveraging popular JupyterLab Notebooks (Python and R), empowers data scientists to analyze aggregated data in a seamless and secure environment (Figure 4).

In the method and algorithm development phase, users can develop or modify pipelines in a sandbox environment. There, they can rapidly build, test, and iterate on machine learning models as needed. Users have access to a broad range of standard libraries, such as TensorFlow[3] or scikit-learn,[4] and can easily bring in their own custom libraries. When users are ready to move to the production phase, ICA enables conversion of notebooks into tools. These tools will then be available in the ICA 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 theratpeutics, and human diagnostics. ICA employs various digital and administrative measures to meet even the most demanding data security requirements:

- Data uploaded from sequencing instruments are encrypted using the AES 256 standard and protected by transfer layer security (TLS)
- Data within ICA 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 practices, [5]
- Authentication service is supported by SAML 2.0 to manage institutional users and passwords (optional)
- Audit reports support traceability of data provenance

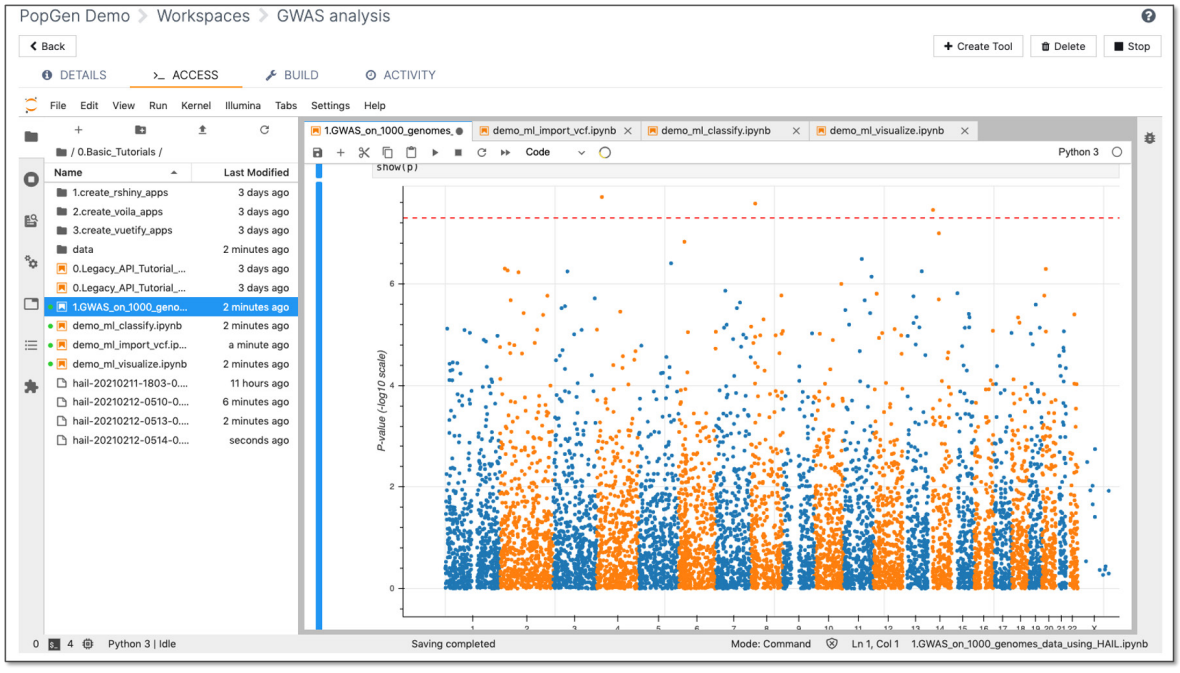

Figure 4: Interactive analysis and visualization—ICA supports use of Jupyter Notebooks for visual exploration of multidimensional data.

ICA also supports customers operating in regulated environments, who must comply with stringent requirements:

- Current data protection laws such as General Data Protection Regulation (GDPR)[6] and Health Insurance Portability and Accountability Act (HIPAA)[7]
- International Organization for Standardization (ISO) 27001 information security management system[8]
- Guaranteed data residency to address local regulatory and compliance requirements

Ordering information

Product	Catalog no.
ICA Professional Annual Subscription	20044876
ICA Enterprise Annual Subscription	20038994
ICA Enterprise Compliance Add-on	20066830
ICA Training and Onboarding	20049422

Learn more

Visit illumina.com/ConnectedAnalytics

References

1. Enabling responsible genomic data sharing for the benefit of human health. Global Alliance for Genomics & Health website. www.ga4gh.org. Accessed October 22, 2020.
2. Illumina DRAGEN Bio-IT Platform | Variant calling & secondary genomic analysis. Illumina website. www.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html. Accessed October 22, 2020.
3. TensorFlow. TensorFlow website. tensorflow.org. Accessed January 11, 2021.
4. scikit-learn: machine learning in Python. scikit-learn website. scikit-learn.org/stable/. Accessed January 11, 2021.
5. Cloud Security—Amazon Web Services (AWS). Amazon website. aws.amazon.com/security. Accessed October 22, 2020.
6. General Data Protection Regulation (GDPR) Compliance Guidelines. GDPR website. gdpr.eu. Accessed January 11, 2021.
7. US Department of Health & Human Services. Health Information Privacy. HHS website. hhs.gov/hipaa/index.html. Accessed January 11, 2021.
8. International Organization for Standardization. ISO-ISO/IEC 27001—Information security management. ISO website. iso.org/isoiec-27001-information-security.html. Accessed January 11, 2021.
9. iCredits for Data Storage and Analysis | Illumina Analytics. Illumina website. www.illumina.com/products/by-type/informatics-products/icredits.html. Accessed October 22, 2020.

illumina®

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

M-GL-00684 v2.0.