

illumina Connected Analytics

Transforme el aluvión
de datos en hallazgos

- Esta solución optimizada, que ofrece resultados a partir de lecturas, impulsa los flujos de trabajo de datos de ómicas a escala
- Las interfaces centradas en el usuario son compatibles con flujos de trabajo personalizados y aprovechan las avanzadas herramientas de ciencia de datos
- Entorno seguro diseñado para favorecer la privacidad de los datos

illumina®

Introducción

Los avances en las tecnologías de secuenciación de nueva generación (NGS, next-generation sequencing) han transformado considerablemente la celeridad con que se lleva a cabo la investigación clínica y en ciencias biológicas. Con el aumento de la velocidad de la secuenciación y la disminución de su coste, la capacidad de generar datos superará con creces la capacidad de extraer información biológica y clínica de ellos. Para poder cumplir los retos de la gestión de datos segura, la colaboración, el análisis de datos complejos y la extracción de conocimientos a partir de datos a escala es necesario tener la capacidad de pasar con facilidad de la generación de datos a su interpretación. Illumina Connected Analytics (ICA) se ha diseñado para gestionar, analizar e interpretar esta enorme cantidad de datos.

ICA es una exhaustiva plataforma de gestión y análisis de datos basada en la nube que permite a los investigadores agregar, explorar y compartir grandes volúmenes de datos multiómicos en un entorno seguro, flexible y adaptable (Figura 1, Tabla 1). ICA ofrece:

- Una integración directa con el flujo de trabajo de generación de datos, incluidos los sistemas de secuenciación de Illumina
- Un potente análisis secundario con la plataforma de tecnología bioinformática DRAGEN™ Bio-IT¹
- Una agregación de datos flexible y un almacenamiento de datos seguro
- Un entorno de ciencia de datos dinámico e interactivo para funciones avanzadas de aprendizaje automático e inteligencia artificial

Flujo de trabajo optimizado

ICA es un componente central para laboratorios en los que se llevan a cabo estudios de NGS con los sistemas de secuenciación de Illumina. Gracias a la flexibilidad de los recursos informáticos que proporciona el entorno informático basado en la nube, ICA permite realizar operaciones a cualquier escala, que abarcan desde el cribado ocasional hasta decenas de miles de células en complejos proyectos de células únicas o incluso la secuenciación del genoma completo a escala poblacional, con la misma arquitectura.

Tabla 1: resumen de la solución ICA

	Característica	Ventaja
Seguridad y privacidad	Cumplimiento normativo	Es conforme con los estándares normativos locales, regionales y globales, los estándares de la Ley HIPAA y el Reglamento General de Protección de Datos (RGPD), y las certificaciones ISO 13485 e ISO 27001
	Controles de seguridad	Mantiene una estricta segregación de datos, cifrado "en tránsito" (TLS 1.2) y "en reposo" (AES 256)
	Registro de auditoría	Mantiene un registro de la actividad para hacer un seguimiento de quién ha accedido a qué datos y cuándo
	Inicio de sesión individual (Single sign-on, SSO) (opcional)	Aprovecha las credenciales institucionales para controlar el acceso
Recursos	Recursos informáticos a demanda	Reducción de los costes, ya que tan solo se paga por los recursos informáticos en el motor de procesamiento
	Escala a demanda	Necesidades de almacenamiento y recursos informáticos a escala basadas en la nube para cumplir el nivel de demanda actual
	Plataforma y panel de control para su uso	Muestra visualmente las solicitudes de recursos para comprender, gestionar y prever las necesidades de manera eficiente
Gestión	Gestión de proyectos y usuarios	Gestione el acceso y la actividad de los usuarios para una privacidad minuciosa
	Uso compartido de datos	Establece comunicación entre almacenamientos de datos para una colaboración global y a gran escala
	Archivo de datos	Reduce costes mediante el archivado de datos no utilizados en niveles de almacenamiento de menor coste
Capacidad de uso	Integración directa de sistemas de secuenciación	Flujo de datos directo desde los sistemas de secuenciación de Illumina
	Generador de procesos visual	Cree procesos sin escribir el código
	Herramientas y procesos	Aproveche los procesos listos para usar y las herramientas de terceros
	API y CLI	Interactúe con la plataforma mediante programación utilizando herramientas basadas en las preferencias del usuario
	Cuenta "Bring your own cloud" (Traiga su propia nube)	Conéctese a su propia nube
	Visualización de datos	Cree diagramas visuales dinámicos y aplicaciones web interactivas para mostrar datos con los paquetes R y Python
Herramientas avanzadas	Soporte de Docker y CWL	Escriba procesos con el estándar Common Workflow Language (CWL) y ejecute análisis en la nube de manera fácil
	API RESTful, conformes con la GA4GH	Permita el acceso mediante programación a herramientas y datos, así como la interoperabilidad con otros entornos de software
	Integrado con JupyterLab	Realice análisis de datos avanzados; cree y entrene modelos de AI/ML (inteligencia artificial y aprendizaje automático) con R y Python
	Agregación y consulta de datos	Haga consultas de datos a nivel poblacional utilizando SQL

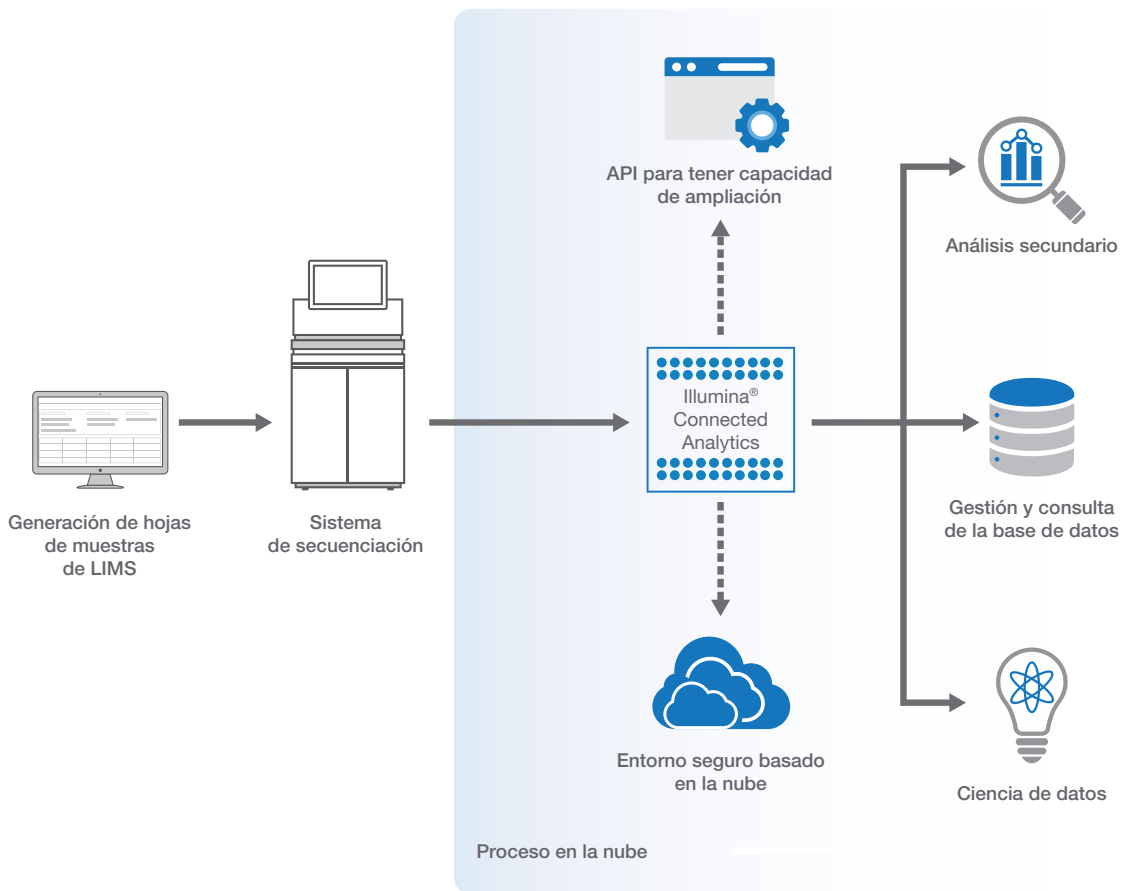


Figura 1: ICA forma la base para la gestión y el análisis de datos

A través de BaseSpace™ Sequence Hub,² los usuarios pueden integrar su plataforma y sus datos de secuenciación directamente con el entorno ICA. Los flujos de trabajo automatizados transmiten los datos del instrumento a la nube en tiempo real a medida que se generan y garantizan la disponibilidad de las lecturas para su análisis en ICA lo más rápido posible.

Una vez que se encuentran en el entorno ICA, los datos se pueden analizar automáticamente con los procesos listos para usar de DRAGEN o bien con procesos personalizados, en función del flujo de trabajo especificado. La amplia variedad de opciones de análisis va desde el control de calidad hasta la agregación de datos y las avanzadas herramientas de ciencia de datos, para un procesamiento de datos rápido y flexible. ICA proporciona una plataforma con capacidad de ampliación con un completo juego de interfaces de programación de aplicaciones (application program interface, API) RESTful y una herramienta de interfaz de la línea de comandos (command-line interface, CLI).

Estas API aumentan al máximo la eficiencia de los flujos de trabajo a medida que se transfieren los datos, se accede a ellos o se utilizan en todo su ciclo de vida; incluyen API que cumplen las guías de la Alianza Global para la Genómica y la Salud (Global Alliance for Genomics and Health, GA4GH).³

Gestión y control de datos

Con el aumento de la generación de datos se acrecienta la necesidad de infraestructuras que permitan compartir, reutilizar e integrar datos dentro de la comunidad científica con el objetivo de amplificar el valor de los conjuntos de datos individuales. Para abordar esta necesidad, la solución ICA incorpora diversas funciones diseñadas para adoptar las prácticas óptimas en la gestión de datos.

Control de acceso

El control de acceso detallado permite al administrador configurar permisos y aprovechar las credenciales institucionales existentes para controlar el acceso. Un registro de auditoría permite monitorizar eventos y cambios mediante el registro de cada usuario cuando accede a la plataforma y de sus acciones durante el uso de la misma. Esto permite imponer el cumplimiento y la rendición de cuentas.

Formato abierto

Con el fin de respaldar un método multiómico para la investigación, la solución ICA se ha diseñado como una plataforma independiente de los datos. De esta manera, permite analizar múltiples tipos de datos, por ejemplo, datos moleculares, clínicos, fenotípicos y desestructurados, como imágenes.

Colaboración

ICA favorece la colaboración sin fronteras geográficas a la vez que se mantiene el cumplimiento. Permite proporcionar y compartir datos y herramientas al instante con otros usuarios, manteniendo siempre la integridad y la privacidad de los datos. Además, los datos y las herramientas de análisis que se alojan en una nube externa de origen se pueden importar al entorno de la solución ICA para analizar y compartir.

Transformación de lecturas en datos

ICA ofrece varias opciones de análisis de datos secundarios, con lo que se optimiza el flujo de trabajo de obtención de resultados a partir de las lecturas. Gracias a la flexibilidad para utilizar procesos listos para usar o construir y configurar procesos personalizados, ICA es compatible con prácticamente todas las aplicaciones informáticas.

Opciones listas para usar

ICA aporta potentes herramientas y procesos para el procesamiento de datos listos para usar, como el acceso a la plataforma de tecnología bioinformática DRAGEN Bio-IT,¹ que permite un análisis secundario rápido y preciso de datos de secuenciación (Figura 2).

Personalización de procesos

Los bioinformáticos pueden importar herramientas existentes de un repositorio de imágenes de Docker, o construir y editar nuevos procesos con el uso del estándar Common Workflow Language (CWL) y el editor gráfico de procesos. Los operadores del laboratorio y otro personal científico también pueden iniciar procesos con facilidad utilizando la interfaz del usuario de diseño intuitivo.

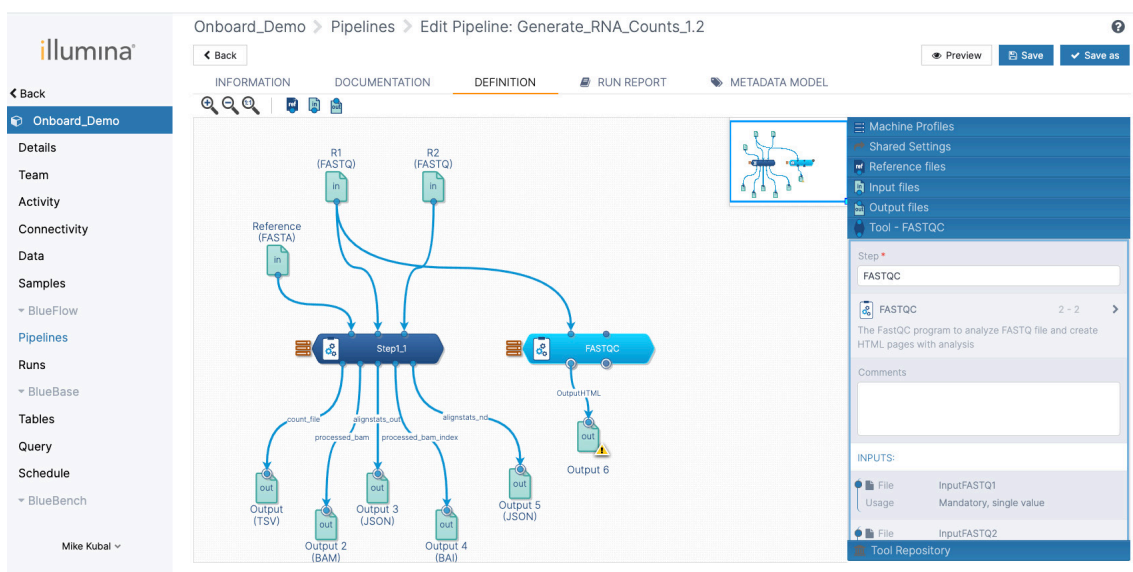


Figura 2: proceso de DRAGEN en ICA. Los usuarios pueden acceder a procesos listos para usar de la plataforma de tecnología bioinformática DRAGEN Bio-IT para obtener un análisis secundario rápido y preciso, que pasa de las lecturas al informe.

Para un desarrollo de procesos acelerado, los usuarios también pueden acceder a las ICA Reference Solutions, una colección de procesos de análisis que se pueden optimizar para adaptarse a las necesidades específicas.

Aprendizaje continuo

La solución ICA automatiza los complejos pasos de agregación e integración para crear un sistema de gestión de conocimientos funcional que engloba datos de millones de muestras (Figura 3). Lo que hace es capturar cualquier tipo de datos disponibles, ya sean genotípicos, fenotípicos, metadatos, anotaciones u otra información asociada. Los usuarios pueden definir sus propios modelos de datos, redactar sus propias consultas y explorar conexiones entre los conjuntos de datos, según sus necesidades. Los datos generados en ICA representan una gran cantidad de información que se puede utilizar para descubrir nuevos biomarcadores, estratificar poblaciones de pacientes, supervisar el rendimiento del ensayo a lo largo del tiempo y mucho más.

Compatible con casi todas las aplicaciones de genómica

Con las innumerables exploraciones de datos en curso, es esencial tener la capacidad de desarrollar y personalizar los algoritmos. Un módulo de programación interactivo, que aprovecha los populares Jupyter Notebooks (Python y R), permite a los científicos de datos analizar datos agregados en un entorno ininterrumpido y seguro (Figura 4).

En la fase de desarrollo del método y el algoritmo, los usuarios pueden desarrollar sus propios procesos o modificar los existentes en un entorno aislado, en el que pueden crear, probar e iterar rápidamente en modelos de aprendizaje automático, según sea necesario. Los usuarios tienen acceso a una amplia variedad de bibliotecas estándar, como TensorFlow⁴ o scikit-learn,⁵ y pueden introducir fácilmente sus propias bibliotecas personalizadas. Cuando los usuarios están listos para pasar a la fase de producción, ICA permite la conversión de los blocs de notas en herramientas. Posteriormente, estas herramientas estarán disponibles en el repositorio de herramientas de ICA y se incorporarán en los procesos de producción.

Figura 3: ICA permite la generación de datos, la minería de datos y el aprendizaje continuo. Los usuarios pueden explorar conexiones entre conjuntos de datos para responder a sus preguntas.

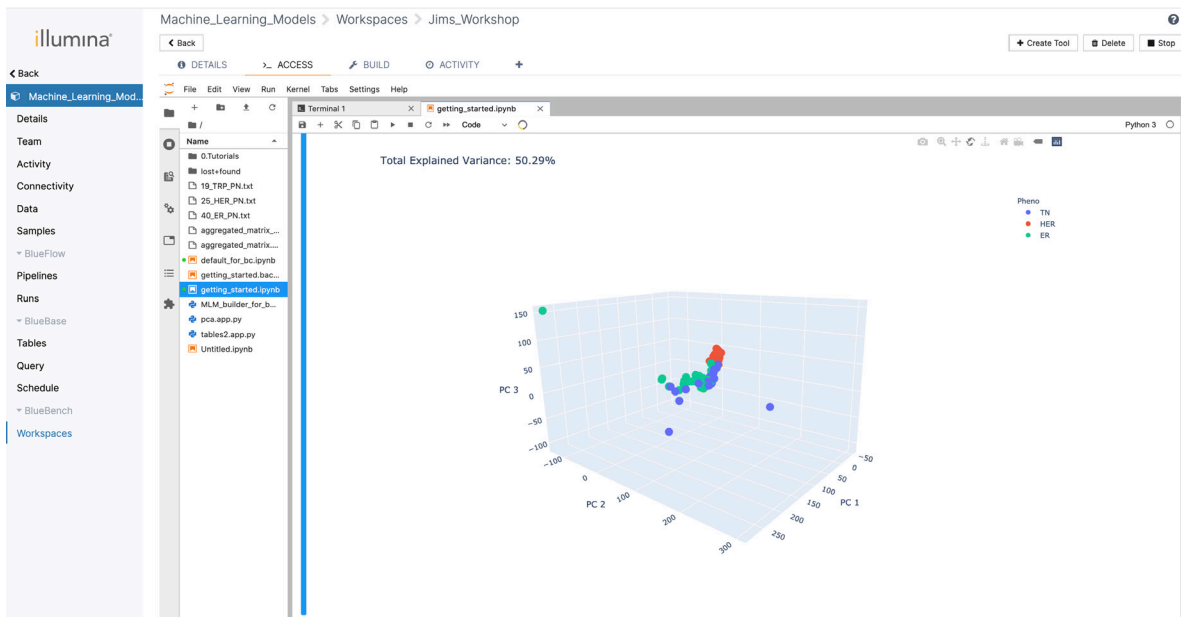


Figura 4: análisis interactivo y visualización. ICA es compatible con el uso de Jupyter Notebook para la exploración visual de datos multidimensionales.

Entorno en el que lo más importante es la seguridad, con apoyo para el cumplimiento

Cuando se operan datos en un entorno basado en la nube, la seguridad es de vital importancia. ICA emplea diversas medidas físicas, electrónicas y administrativas para cumplir incluso los requisitos de seguridad de datos más exigentes:

- Los datos que se cargan desde instrumentos de secuenciación se cifran utilizando el estándar AES 256 y se protegen mediante el protocolo de seguridad de la capa de transporte TLS (transfer layer security)
- Los datos dentro de ICA se alojan en Amazon Web Services (AWS), y aprovechan las prácticas óptimas de AWS Well-Architected, que cumplen una gran variedad de normas de seguridad aceptadas por el sector⁶

- El servicio de autenticación es compatible con SAML 2.0 para gestionar los usuarios y las contraseñas institucionales (opcional)
- Informes de auditoría para la trazabilidad de la procedencia de los datos

La solución ICA también pueden utilizarla los clientes que operan en entornos regulados y que deben cumplir estrictos requisitos:

- Legislación actual relativa a la protección de datos, como el Reglamento General de Protección de Datos (RGPD)⁷ y la Ley de portabilidad y responsabilidad del seguro médico (Health Insurance Portability and Accountability Act, HIPAA)⁸
- Norma de la Organización Internacional de Normalización (International Organization for Standardization, ISO) 13485 relativa a los Sistemas de gestión de la calidad⁹ y norma ISO 27001 relativa a los Sistemas de gestión de la seguridad de la información¹⁰
- Residencia de datos garantizada para asegurar el gestión de los requisitos normativos y de cumplimiento locales

Opciones flexibles

Para mayor flexibilidad de adquisición, ICA está disponible en forma de suscripción anual. Para la facturación se emplean iCredits en función del uso de herramientas y almacenamiento.¹¹ Los iCredits pueden adquirirse por adelantado o bien se facturan mensualmente.

Estudios multiómicos flexibles

A medida que aumenta la rapidez de la generación de datos de NGS y disminuye su coste, las plataformas de datos avanzadas que permiten a los investigadores obtener informes fácilmente y a escala a partir de las lecturas se convierten en una parte crucial del éxito. Gracias a soluciones potentes que respaldan la colaboración global mediante un acceso centralizado a los datos distribuidos, a los procesos personalizables y a los procesos listos para usar, a herramientas de ciencia de datos y a un entorno seguro conforme a las normativas internacionales, ICA ofrece a los usuarios la posibilidad de aprovechar todo el potencial de los datos multiómicos.

Información adicional

Visite la página www.illumina.com/ConnectedAnalytics

Referencias

1. Illumina DRAGEN Bio-IT Platform | Variant calling & secondary genomic analysis. Sitio web de Illumina. www.illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html. Acceso: 22 de octubre de 2020.
2. BaseSpace Sequence Hub | Cloud-based genomics computing. Sitio web de Illumina. www.illumina.com/basespace. Acceso: 11 de enero de 2021.
3. Enabling responsible genomic data sharing for the benefit of human health. Sitio web de Global Alliance for Genomics & Health. www.ga4gh.org. Acceso: 22 de octubre de 2020.
4. TensorFlow. Sitio web de TensorFlow. tensorflow.org. Acceso: 11 de enero de 2021.
5. scikit-learn: machine learning in Python. Sitio web de scikit-learn. scikit-learn.org/stable/. Acceso: 11 de enero de 2021.
6. Cloud Security—Amazon Web Services (AWS). Sitio web de Amazon. aws.amazon.com/security. Acceso: 22 de octubre de 2020.
7. General Data Protection Regulation (GDPR) Compliance Guidelines. Sitio web RGPD. gdpr.eu. Acceso: 11 de enero de 2021.
8. US Department of Health & Human Services. Health Information Privacy. Sitio web de HHS. hhs.gov/hipaa/index.html. Acceso: 11 de enero de 2021.
9. International Organization for Standardization. ISO-ISO 13485:2016-Medical devices—Quality management systems—Requirements for regulatory purposes. Sitio web ISO. iso.org/standard/59752.html. Acceso: 11 de enero de 2021.
10. International Organization for Standardization. ISO-ISO/IEC 27001—Information security management. Sitio web ISO. iso.org/isoiec-27001-information-security.html. Acceso: 11 de enero de 2021.
11. iCredits for Data Storage and Analysis | Illumina Analytics. Sitio web de Illumina. www.illumina.com/products/by-type/informatics-products/icredits.html. Acceso: 22 de octubre de 2020.

Datos para realizar pedidos

Producto	N.º de catálogo
ICA Enterprise	20038994
ICA Data Science	20044877
Illumina Analytics: 1 iCredit	20042038
Illumina Analytics: 1000 iCredits	20042039
Illumina Analytics: 5000 iCredits	20042040
Illumina Analytics: 50 000 iCredits	20042041
Illumina Analytics: 100 000 iCredits	20042042
Facturación por consumo	20012931

a. No incluye funciones de ciencia de datos.
b. Proporciona acceso a Notebooks (Jupyter, R) y al marco de AI/ML.
c. El importe indicado en el presupuesto representa el importe que no se debe superar. A los clientes se les facturará mensualmente por el consumo de recursos informáticos, almacenamiento y aplicaciones de terceros hasta la cantidad asociada al n.º de catálogo 20012931.

illumina®

1.800.809.4566 (llamada gratuita, EE. UU.) | tel.: +1.858.202.4566
techsupport@illumina.com | www.illumina.com

© 2020 Illumina, Inc. Todos los derechos reservados. Todas las marcas comerciales pertenecen a Illumina, Inc. o a sus respectivos propietarios. Si desea consultar información específica sobre las marcas comerciales, visite www.illumina.com/company/legal.html.
N.º de pub. 986-2020-009-A ESP. QB11606.