

Illumina Connected Analytics

Per trasformare le procedure operative della bioinformatica

- Integrazione semplificata con il sistema di sequenziamento
- Produttività scalabile per soddisfare le esigenze di elaborazione dei campioni
- Interfaccia intuitiva per selezionare, creare e personalizzare i flussi di lavoro

Introduzione

I progressi compiuti nell'ambito delle tecnologie di sequenziamento di nuova generazione (NGS, next-generation sequencing) hanno drasticamente cambiato il ritmo con il quale evolvono scienze biologiche e ricerca clinica. Man mano che la velocità di sequenziamento aumenta e i costi si riducono, la capacità di generare dati supera di gran lunga la possibilità di estrarre informazioni biologiche e cliniche dai dati stessi. Per affrontare le complessità dello sviluppo e dell'implementazione delle pipeline, scalare i flussi di lavoro informatici e garantire una gestione sicura dei dati, è necessaria una piattaforma flessibile e completa. Illumina Connected Analytics consente di creare, modificare e distribuire pipeline analitiche flessibili garantendo al contempo la privacy, la sicurezza e la conformità dei dati.

Illumina Connected Analytics è una piattaforma per dati bioinformatici sicura e basata sul cloud, che consente di utilizzare le risorse informatiche per ottenere informazioni scientifiche (Figura 1, Tabella 1). Illumina Connected Analytics permette ai clienti di creare flussi di lavoro:

- connessi e più facilmente integrabili con sistemi e software Illumina;
- scalabili, con la possibilità di gestire, analizzare e interrogare enormi set di dati;
- flessibili, realizzabili e personalizzabili dagli utenti mediante le applicazioni DRAGEN™ e pipeline di analisi personalizzate.

Flusso di lavoro ottimizzato

Illumina Connected Analytics è un componente fondamentale per i laboratori che conducono studi NGS con sistemi di sequenziamento Illumina. Sfruttando la flessibilità delle risorse resa possibile dal cloud computing, Illumina Connected Analytics supporta con la stessa architettura operazioni di qualsiasi portata, dallo screening occasionale a complessi progetti a singola cellula che coinvolgono decine di migliaia di cellule, fino al sequenziamento dell'intero genoma a livello di popolazione. Gli utenti possono integrare gli strumenti esistenti con Illumina Connected Analytics senza alcuna difficoltà.

Sulla piattaforma Illumina Connected Analytics, i dati possono essere analizzati automaticamente con pipeline di analisi secondaria DRAGEN pronte all'uso o personalizzate, a seconda del flusso di lavoro specificato. L'ampia gamma di opzioni di analisi spazia dal controllo qualità all'aggregazione dei dati, con strumenti di data science avanzati per un'elaborazione dei dati rapida e scalabile. Illumina Connected Analytics offre una piattaforma ampliabile con un ricco set di interfacce di programmi applicativi (API, application program interface) RESTful e uno strumento di

interfaccia a riga di comando (CLI, command-line interface). Queste API ottimizzano l'efficienza dei flussi di lavoro mentre si visualizzano, si trasferiscono e si utilizzano i dati nel corso del loro intero ciclo di vita.

Tabella 1: Illumina Connected Analytics in sintesi

	Caratteristica	Vantaggio
Sicurezza e privacy	Conformità	Prodotto sviluppato nell'ambito del sistema di gestione della qualità (QMS, Quality Management System) di Illumina in conformità agli standard normativi locali, regionali e globali, tra cui GDPR, ² HIPAA, ³ ISO 27001/27701, ^{4,5} e ISO 13485 ⁶
	Controlli di sicurezza	Rigida separazione dei dati, con crittografia in transito (TLS 1.2) e a riposo (AES 256)
	Audit trail	Costante monitoraggio delle attività e del registro eventi per il controllo degli accessi
	Autenticazione a più fattori (MFA, multifactor authentication)	Applicazione di credenziali istituzionali per la gestione e il controllo degli accessi al sistema
Risorse	Scalabilità su richiesta	Scalabilità in base alle esigenze di calcolo e di archiviazione su cloud per soddisfare i livelli variabili della domanda
	Risorse di calcolo su richiesta	Riduzione dei costi, poiché si paga per le sole risorse di calcolo nel motore della pipeline
Gestione	Gestione di progetti e utenti	Gestione di accessi e attività degli utenti per la privacy granulare
	Condivisione dei dati	Riduzione dei silos di dati per promuovere la collaborazione globale e favorire la distribuzione dei dati
	Archiviazione dei dati	Riduzione dei costi per la gestione a lungo termine dei dati con diversi livelli di archiviazione
	Utilizzo del proprio bucket su cloud AWS	Accesso ai dati archiviati in account cloud AWS gestiti privatamente
Usabilità e accessibilità	Integrazione del sistema di sequenziamento	Trasmissione fluida dei dati dai sistemi di sequenziamento Illumina alle risorse di archiviazione e analisi di Illumina Connected Analytics
	Strumenti e pipeline	Applicazione di pipeline DRAGEN pronte all'uso o importazione di strumenti esistenti
	API e CLI	Possibilità di interagire in modo programmatico con la piattaforma, gli strumenti e i dati tramite API RESTful o CLI
	Interfaccia grafica intuitiva	Accesso a dati e strumenti tramite un'interfaccia web grafica; visualizzazione dei dati con i pacchetti R e Python
Strumenti avanzati	Compatibilità con Docker, Nextflow e CWL	Scrittura di pipeline nel Common Workflow Language e semplice avvio delle analisi nel cloud
	Integrazione con JupyterLab	Esecuzione di script personalizzati, creazione e addestramento di modelli di intelligenza artificiale/apprendimento automatico e interazione con i dati sulla piattaforma tramite ambienti notebook collaborativi
	Aggregazione e interrogazione dei dati	Organizzazione e interrogazione di set di dati multiomici strutturati per consentire l'analisi terziaria

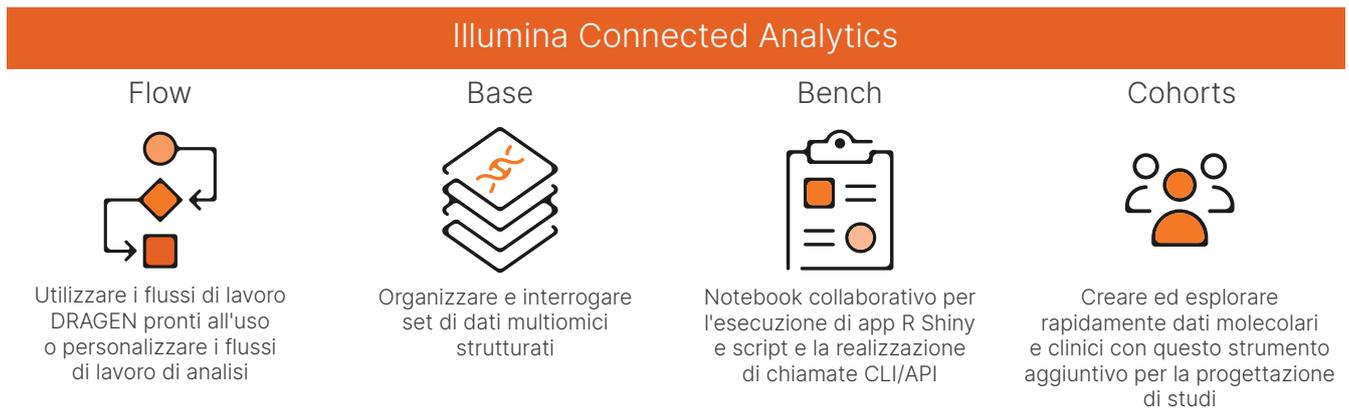


Figura 1: componenti modulari di Illumina Connected Analytics. Elaborano, aggregano e analizzano i dati in un ambiente di gestione dei dati sicuro e con controllo degli accessi.

BaseSpace™ Sequence Hub

Tutti gli abbonamenti a Illumina Connected Analytics includono BaseSpace Sequence Hub, un'estensione diretta degli strumenti Illumina. L'integrazione di BaseSpace Sequence Hub consente di aumentare l'efficienza del laboratorio grazie alla possibilità di impostare le corse, monitorare la qualità di queste ultime e trasmettere i dati di sequenziamento direttamente sul cloud. I dati codificati vengono trasferiti dallo strumento a BaseSpace Sequence Hub e un set di applicazioni selezionate all'interno di un ambiente sicuro semplifica l'analisi e la gestione dei dati.

Dati fruibili ricavati dalle letture

Illumina Connected Analytics offre varie opzioni per l'analisi secondaria dei dati, semplificando il flusso di lavoro letture-risultati. La possibilità di usare pipeline già pronte o di creare e configurare pipeline personalizzate assicura la compatibilità di Illumina Connected Analytics praticamente con qualsiasi applicazione informatica.

Opzioni pronte all'uso

Illumina Connected Analytics fornisce potenti strumenti e pipeline pronti all'uso per l'elaborazione dei dati, incluso l'accesso all'analisi secondaria DRAGEN,¹ che esamina i dati di sequenziamento in modo accurato, efficiente e completo (Figura 2).

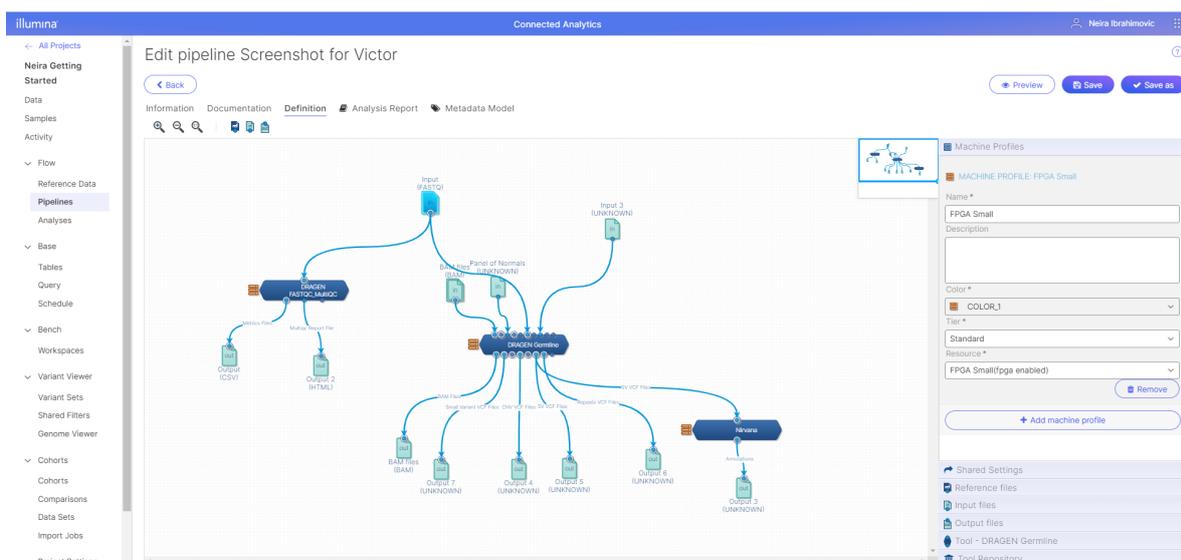


Figura 2: Illumina Connected Analytics Flow. Creazione, distribuzione e versioning flessibili delle pipeline analitiche.

Personalizzazione delle pipeline

I bioinformatici possono importare strumenti di analisi esistenti da un repository di immagini Docker Hub⁷ oppure creare e modificare nuove pipeline usando Nextflow, il Common Workflow Language (CWL) e l'editor grafico di pipeline. Inoltre, gli operatori di laboratorio e altri tecnici possono avviare pipeline in tutta semplicità utilizzando l'interfaccia utente dal design intuitivo.

Gestione e controllo dei dati

L'aumento dei volumi di dati generati implica una maggiore necessità di infrastrutture che ne consentano la condivisione, il riutilizzo e l'integrazione all'interno della comunità scientifica, in modo da amplificare il valore dei singoli set di informazioni. Per soddisfare questa esigenza, Illumina Connected Analytics incorpora diverse caratteristiche progettate proprio per consentire l'adozione delle migliori pratiche nella gestione dei dati.

Controllo degli accessi

Il controllo granulare degli accessi permette agli amministratori di impostare autorizzazioni e utilizzare le credenziali istituzionali esistenti per controllare gli accessi. Il registro di controllo riporta tutti gli eventi e le modifiche registrando ogni accesso alla piattaforma e le operazioni eseguite dagli utenti durante l'uso. In questo modo, si garantiscono conformità e responsabilità.

Formato aperto

Illumina Connected Analytics è progettato come piattaforma indipendente dai dati. Supporta, infatti, l'analisi di diversi tipi di dati, quali dati molecolari, clinici, fenotipici e non strutturati (come le immagini).

Collaborazione

Illumina Connected Analytics permette una collaborazione senza confini geografici, sempre attenta alla conformità. Dati e strumenti possono essere forniti e condivisi immediatamente con altri utenti in modo da preservare l'integrità dei dati e la privacy. Inoltre, i dati e gli strumenti analitici ospitati in origini cloud esterne possono essere importati in Illumina Connected Analytics per l'analisi e la condivisione.

Aggregazione e interrogazione dei dati

Illumina Connected Analytics automatizza procedure complesse di aggregazione e integrazione per creare un sistema funzionale di gestione delle conoscenze che comprende dati provenienti da milioni di campioni. Cattura pressoché tutti i tipi di dati disponibili: genotipici, fenotipici, metadati, annotazioni e altre informazioni associate. Gli utenti possono definire i propri modelli di dati, scrivere query ed esplorare le connessioni tra i set di informazioni in base alle esigenze specifiche. I dati aggregati in Illumina Connected Analytics rappresentano una miniera di informazioni da utilizzare per individuare nuovi biomarcatori, stratificare le popolazioni di pazienti, monitorare le prestazioni dei saggi nel tempo e molto altro ancora.

Uso delle coorti per l'esplorazione dei dati clinico-genomici su larga scala

Illumina Connected Analytics Cohorts è uno strumento per la progettazione di studi che consente di creare ed esplorare rapidamente le coorti (Figura 3). Illumina Connected Analytics Cohorts accelera il processo che dalla scoperta genomica porta a studi significativi e consente un migliore processo decisionale in merito alla progettazione dello studio grazie all'aggregazione di dati multiomici e fenotipici in un unico luogo.

Integra le coorti grazie all'inclusione di set di dati pubblici armonizzati, tra cui The Cancer Genome Atlas (TCGA), Broad Rare Genomes Project, 1000 Genomes Project e Gene Expression Omnibus. Consente di formulare ipotesi in pochi minuti anziché in settimane, sfruttando un browser di coorti integrato. Offre l'accesso a set di dati pubblici accuratamente selezionati. Mette a disposizione semplici strumenti visivi per selezionare e visualizzare le coorti, senza la necessità di interrogazioni complesse.

Ambiente notebook sicuro per l'acquisizione di informazioni

Lo sviluppo e la personalizzazione dell'algoritmo sono componenti essenziali di Illumina Connected Analytics che consentono l'esplorazione approfondita dei dati. Un modulo di programmazione interattiva, che sfrutta i popolari notebook Jupyter, Python ed R, permette ai data scientist di analizzare i dati aggregati in un ambiente sicuro e completamente integrato (Figura 4).



Figura 3: Illumina Connected Analytics Cohorts. Modulo aggiuntivo di Illumina Connected Analytics per la creazione e l'esplorazione rapide dei dati molecolari e clinici.

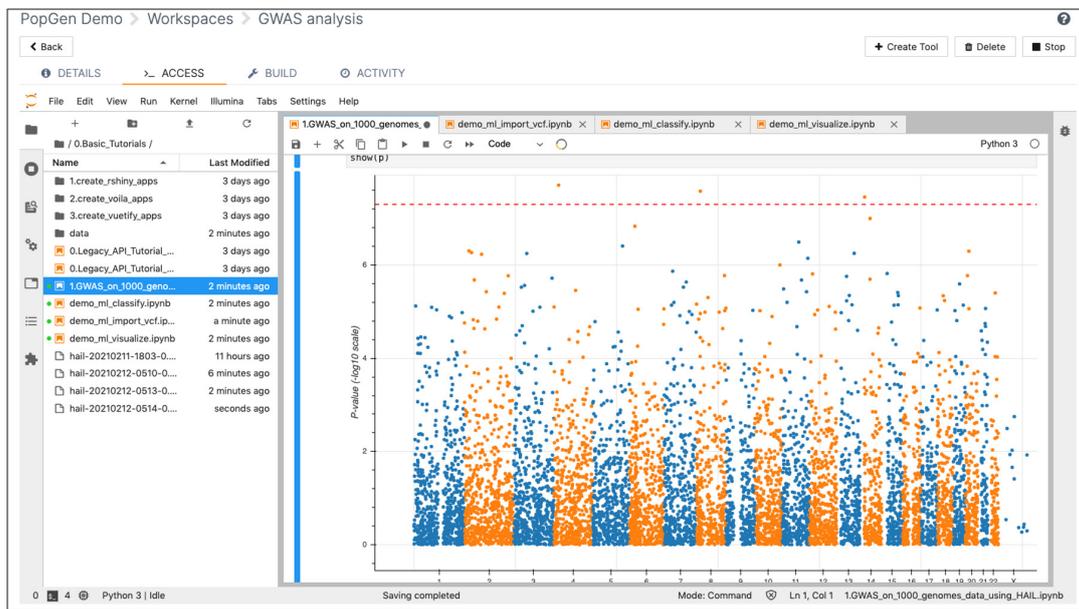


Figura 4: Illumina Connected Analytics Bench. Illumina Connected Analytics Bench integra strumenti di calcolo flessibili e interattivi, inclusi i notebook Jupyter, Python ed R.

Nella fase di sviluppo del metodo e dell'algoritmo, gli utenti possono sviluppare o modificare le pipeline in un ambiente sandbox. Qui hanno accesso a librerie standard e possono facilmente inserire le proprie librerie personalizzate, come TensorFlow⁸ o scikit-learn,⁹ per creare script personalizzati e complessi per l'elaborazione collaborativa dei dati in Illumina Connected Analytics. Quando gli utenti sono pronti a passare alla fase di produzione, Illumina Connected Analytics permette la conversione dei notebook in strumenti, che saranno poi disponibili nel repository degli strumenti di Illumina Connected Analytics e integrati nelle pipeline di produzione.

Sicurezza e conformità al primo posto

La sicurezza è di primaria importanza quando si opera con dati genomici per la ricerca, la terapia clinica e la diagnostica umana. Illumina Connected Analytics impiega diverse misure digitali e amministrative per soddisfare anche i requisiti più rigidi in fatto di sicurezza dei dati.

- I dati caricati dagli strumenti di sequenziamento sono codificati in base allo standard AES256¹⁰ e sono protetti dal protocollo TLS (Transfer Layer Security).
- I dati gestiti da Illumina Connected Analytics sono in hosting sui sistemi Amazon Web Services (AWS) per garantire, tramite le migliori pratiche AWS Well-Architected, la conformità a un'ampia gamma di standard di sicurezza accettati nel settore.¹¹
- Illumina Connected Analytics è in hosting in più di 10 aree geografiche AWS in tutto il mondo per assicurare la conformità alle normative locali e gestire i dati genomici in particolari territori.
- Il servizio di autenticazione è supportato dal Security Assertion Markup Language (SAML) 2.0 per la gestione di utenti e password istituzionali (opzionale).
- I report di audit permettono la tracciabilità della provenienza dei dati.

Illumina Connected Analytics è utilizzabile anche dai clienti che operano in ambienti regolamentati e sono quindi soggetti a requisiti rigorosi. Illumina Connected Analytics è stato sviluppato in conformità al ciclo di vita dello sviluppo software Illumina nell'ambito del sistema di gestione della qualità (QMS, Quality Management System) di Illumina. Inoltre, i processi che rientrano nell'ambito del QMS di Illumina seguono le migliori pratiche del settore e gli standard pertinenti, tra cui:

- Sistema di gestione della sicurezza delle informazioni ISO 27001:2013⁴ e tecniche di sicurezza ISO 27701:2019⁵ dell'Organizzazione internazionale per la normazione (ISO, International Organization for Standardization)

- Migliori pratiche organizzative per il sistema di gestione della qualità (QMS, Quality Management System) ISO 13485⁶
- Regolamento generale sulla protezione dei dati (GDPR, General Data Protection Regulation)²
- Health Insurance Portability and Accountability Act (HIPAA)³
- Garanzia di residenza dei dati per soddisfare i requisiti normativi e di conformità locali

Informazioni per gli ordini

Prodotto	N. di catalogo
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 Svc and Compliance Add-on	20066830

Abbreviazioni: ICA, Illumina Connected Analytics

Maggiori informazioni

[Illumina Connected Analytics](#)

Bibliografia

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



Numero verde 1.800.809.4566 (U.S.A.) | Tel. +1.858.202.4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Tutti i diritti riservati. Tutti i marchi di fabbrica sono di proprietà di Illumina, Inc. o dei rispettivi proprietari. Per informazioni specifiche sui marchi di fabbrica, visitare la pagina web www.illumina.com/company/legal.html.
M-GL-00684 ITA v3.0.