

# Illumina Connected Analytics

Per trasformare l'aumento  
dei dati in nuove scoperte

- Soluzione ottimizzata, con un flusso letture-risultati, che alimenta i flussi di lavoro dei dati -omici su larga scala
- Interfacce incentrate sull'utente che supportano flussi di lavoro personalizzati e sfruttano strumenti avanzati di data science
- Ambiente sicuro costruito per garantire la privacy dei dati

illumina®

## Introduzione

I progressi nelle tecnologie di sequenziamento di futura generazione (NGS) hanno drasticamente cambiato il ritmo con il quale si evolvono scienze biologiche e ricerca clinica. Man mano che la velocità di sequenziamento aumenta e il costo diminuisce, si assiste a un'imponente generazione di dati, che supera di gran lunga la capacità di estrarre informazioni biologiche e cliniche dai dati stessi. Per affrontare le sfide della gestione sicura dei dati, della collaborazione, dell'analisi dei dati complessi e dell'estrazione di conoscenze dai dati su larga scala, si deve poter passare facilmente dalla generazione dei dati alla loro interpretazione. Illumina Connected Analytics (ICA) è una piattaforma pensata per la gestione, l'analisi e l'interpretazione di questa imponente quantità di dati.

Consente l'analisi e la gestione dei dati sul cloud e permette ai ricercatori di aggregare, esplorare e condividere grandi volumi di dati multiomici in un ambiente sicuro, scalabile e flessibile (Figura 1, Tabella 1). La piattaforma ICA offre:

- Integrazione diretta con il flusso di lavoro di generazione dei dati, compresi i sistemi di sequenziamento Illumina
- Potente analisi secondaria con piattaforma DRAGEN™ Bio-IT<sup>1</sup>
- Aggregazione scalabile e archiviazione sicura dei dati
- Ambiente data science dinamico e interattivo per l'apprendimento automatico avanzato e l'intelligenza artificiale

## Flusso di lavoro ottimizzato

La piattaforma ICA è un elemento importante per i laboratori che conducono ricerche NGS con sistemi di sequenziamento Illumina. Sfruttando l'elasticità delle risorse di calcolo offerte dal cloud computing, la piattaforma ICA supporta operazioni scalabili con la stessa architettura, dallo screening occasionale a decine di migliaia di cellule in progetti complessi su singola cellula, fino al sequenziamento dell'intero genoma su scala di popolazione. Grazie a BaseSpace™ Sequence Hub,<sup>2</sup> è possibile integrare piattaforma di sequenziamento e dati direttamente nell'ambiente ICA.

Tabella 1. ICA in sintesi

	Caratteristiche	Vantaggi
Sicurezza e privacy	Conformità	Conformità con le norme regionali, nazionali e internazionali e con gli standard HIPAA e GDPR e le certificazioni ISO13485 e ISO27001
	Controlli di sicurezza	Rigida separazione dei dati, con crittografia "in movimento" (TLS 1.2) e "a riposo" (AES 256)
	Audit trail	Tracciabilità del registro delle attività, per sapere chi ha avuto accesso a quali dati e quando
	Single Sign-On (SSO) (facoltativo)	Sfruttamento delle credenziali istituzionali per controllare l'accesso
Risorse	Risorse di calcolo su richiesta	Riduzione dei costi pagando solo le risorse di calcolo nel motore della pipeline
	Scalabilità su richiesta	Scalabilità di archiviazione su cloud e calcolo delle esigenze per soddisfare i livelli attuali di domanda
	Piattaforma e dashboard di utilizzo	Visualizzazione delle richieste di risorse per comprendere, gestire e anticipare i bisogni in modo efficiente
Gestione	Gestione di progetti e utenti	Gestione dell'accesso e dell'attività degli utenti per una privacy di tipo granulare
	Condivisione dei dati	Collegamento di silos di dati per collaborazioni su larga scala
	Archiviazione dei dati	Riduzione dei costi tramite archiviazione dei dati non utilizzati in livelli di archiviazione con costi inferiori
Usabilità	Integrazione diretta del sistema di sequenziamento	Dati di flusso direttamente dai sistemi di sequenziamento Illumina
	Creatore visivo di pipeline	Creazione di pipeline senza scrivere codice
	Strumenti e pipeline	Sfruttamento immediato delle pipeline e di strumenti di terze parti
	API e CLI	Interazione programmatica con la piattaforma grazie all'utilizzo di strumenti basati sulle preferenze degli utenti
	Account "Bring Your Own Cloud"	Collegamento al proprio cloud privato
	Visualizzazione dei dati	Creazione di grafici visivi dinamici e applicazioni web interattive per visualizzare i dati con i pacchetti R e Python
Strumenti avanzati	Supporto Docker e CWL	Scrittura di pipeline nel linguaggio comune dei flussi di lavoro e lancio facilitato di analisi nel cloud
	API RESTful, conformi a GA4GH	Accesso programmatico a strumenti e dati e interoperabilità con altri ambienti software
	Integrazione con JupyterLab	Esecuzione di analisi avanzate dei dati, costruzione e addestramento di modelli AI/ML con R e Python
	Aggregazione dei dati e query	Esecuzione di query di dati a livello di popolazione usando SQL

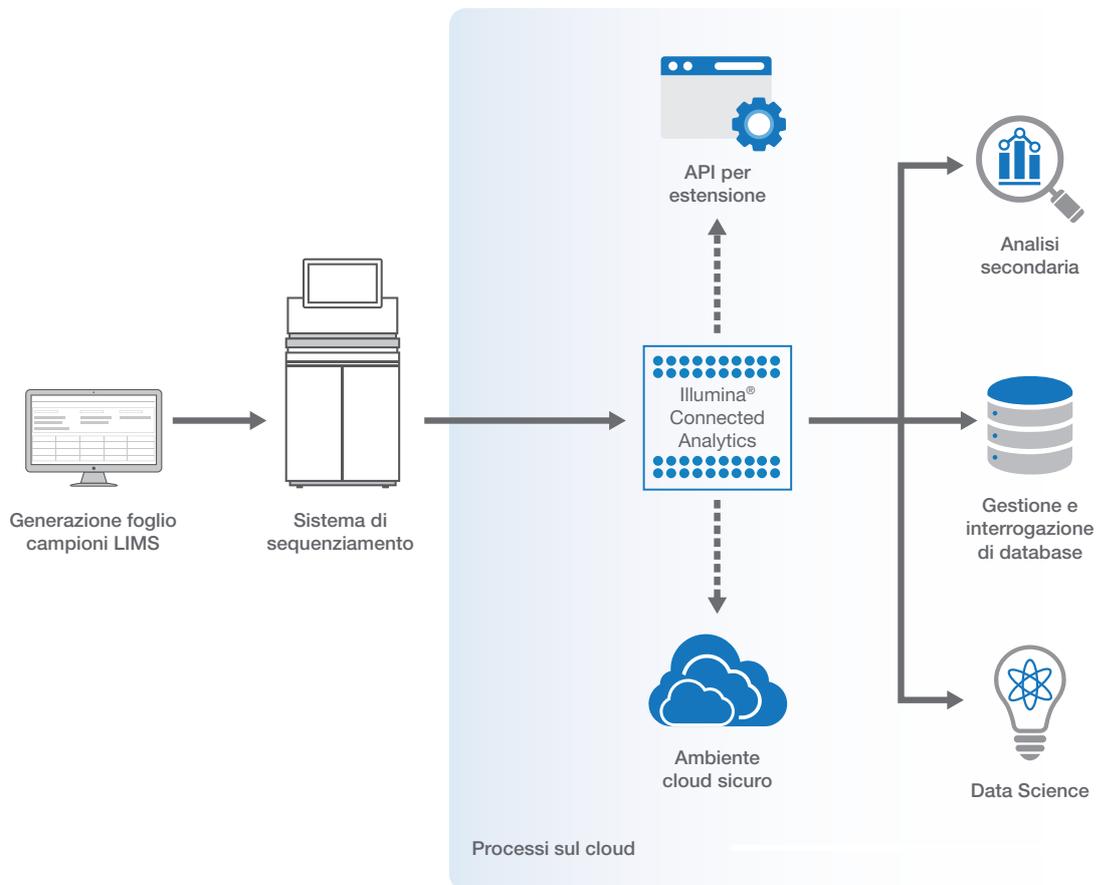


Figura 1. La piattaforma ICA costituisce la base per la gestione e l'analisi dei dati

I flussi di lavoro automatizzati trasmettono i dati dallo strumento al cloud in tempo reale man mano che vengono generati e assicurano la disponibilità delle letture per l'analisi nella piattaforma ICA il più rapidamente possibile.

Una volta entrati nell'ambiente ICA, i dati possono essere analizzati automaticamente con pipeline DRAGEN pronte all'uso o pipeline personalizzate, a seconda del flusso di lavoro specificato. L'ampia gamma di opzioni di analisi spazia dal controllo di qualità all'aggregazione dei dati, con strumenti avanzati di data science per un'elaborazione dei dati rapida e scalabile. Il sistema ICA offre una piattaforma ampliabile con un ricco set di API RESTful e uno strumento di interfaccia a riga di comando (CLI). Queste API, comprese le API conformi alla Global Alliance for Genomics and Health (GA4GH), ottimizzano l'efficienza dei flussi di lavoro mentre si visualizzano, si trasferiscono e si utilizzano i dati nel corso di tutto il ciclo di attività.<sup>3</sup>

## Gestione e controllo dei dati

L'aumento della generazione di dati implica una maggiore necessità di infrastrutture per supportare la condivisione, il riutilizzo e l'integrazione dei dati all'interno della comunità scientifica, in modo da amplificare il valore dei singoli set di dati. Per rispondere a questo bisogno, la piattaforma ICA incorpora alcune caratteristiche pensate proprio per consentire l'adozione delle migliori pratiche nella gestione dei dati.

### Controllo degli accessi

Il controllo granulare degli accessi permette a un amministratore di impostare le autorizzazioni e di sfruttare le credenziali istituzionali esistenti per controllare l'accesso.

Un registro di controllo permette di registrare gli eventi e le modifiche: quando ogni utente accede alla piattaforma e quali azioni esegue mentre la usa. In questo modo, si garantiscono conformità e responsabilità.

## Formato aperto

Per supportare un approccio multiomico alla ricerca, la piattaforma ICA è stata progettata come una piattaforma data-agnostic. Supporta l'analisi di diversi tipi di dati, compresi i dati molecolari, clinici, fenotipici e non strutturati (come le immagini).

## Collaborazione

La piattaforma ICA permette una collaborazione senza confini geografici, sempre preservando la conformità. Dati e strumenti possono essere consegnati istantaneamente e condivisi con altri utenti in modo da preservare l'integrità dei dati e la privacy. Inoltre, i dati e gli strumenti analitici ospitati in un'origine cloud esterna possono essere importati nell'ambiente ICA per l'analisi e la condivisione.

## Dalle letture ai dati

La piattaforma ICA offre varie opzioni per l'analisi secondaria dei dati, ottimizzando il flusso di lavoro letture-risultati. Grazie alla possibilità di usare pipeline già pronte o di costruire e configurare pipeline personalizzate, si può supportare virtualmente qualsiasi applicazione informatica.

## Opzioni pronte all'uso

La piattaforma ICA offre strumenti e pipeline immediati e potenti per l'elaborazione dei dati, compreso l'accesso alla piattaforma DRAGEN Bio-IT,<sup>1</sup> che esegue un'analisi secondaria dei dati di sequenziamento veloce e accurata (Figura 2).

## Personalizzazione delle pipeline

I bioinformatici possono importare strumenti esistenti da un repository di immagini docker oppure costruire e modificare nuove pipeline usando il Common Workflow Language (CWL) e l'editor grafico di pipeline.

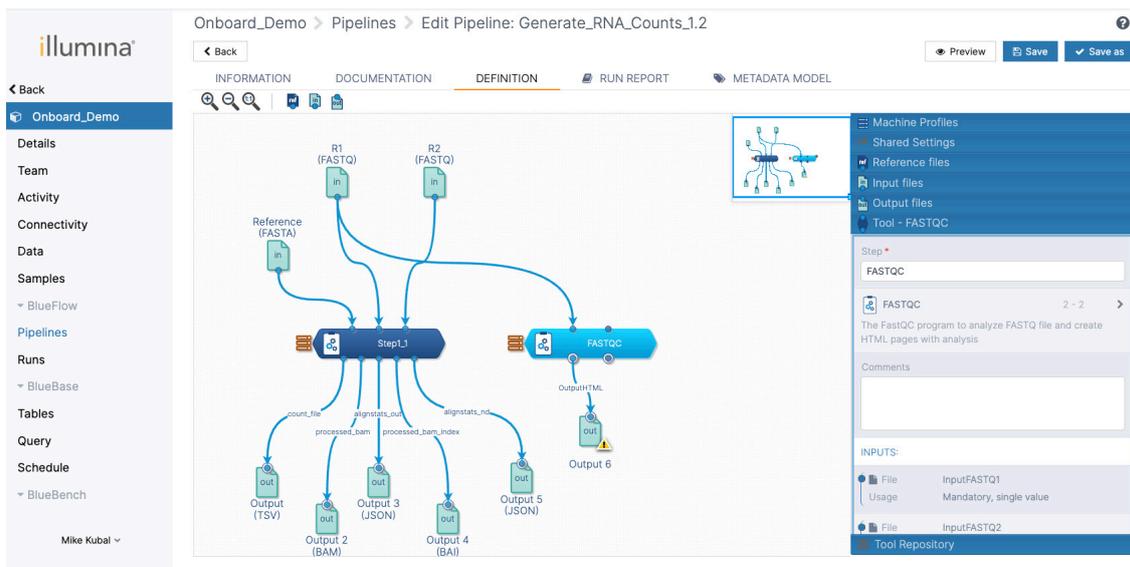


Figura 2. Pipeline DRAGEN nella piattaforma ICA: gli utenti possono accedere a pipeline pronte all'uso dalla piattaforma DRAGEN Bio-IT per analisi secondarie veloci e accurate, con un flusso letture-report.

Inoltre, gli operatori di laboratorio e altri utenti possono lanciare pipeline in modo immediato utilizzando l'interfaccia utente dal design intuitivo. Per accelerare lo sviluppo delle pipeline, gli utenti possono anche accedere alle ICA Reference Solutions, una raccolta di pipeline di analisi ulteriormente ottimizzabile per soddisfare esigenze specifiche.

## Apprendimento continuo

La piattaforma ICA automatizza fasi complesse di aggregazione e integrazione per creare un sistema funzionale di gestione della conoscenza che comprende dati provenienti da milioni di campioni (Figura 3). Cattura tutti i tipi di dati disponibili: genotipici, fenotipici, metadati, annotazioni e altre informazioni associate. Gli utenti possono definire i propri modelli di dati, scrivere le proprie query ed esplorare le connessioni tra i set di dati in base alle loro necessità. I dati aggregati sulla piattaforma ICA rappresentano una miniera di informazioni da utilizzare per scoprire nuovi biomarcatori, stratificare le popolazioni di pazienti, monitorare le prestazioni dei saggi nel tempo e molto altro ancora.

## Possibilità di supportare virtualmente qualsiasi applicazione genomica

Con la miriade di esplorazioni di dati in corso, la capacità di sviluppare e personalizzare gli algoritmi è essenziale. Un modulo di programmazione interattiva, che sfrutta i popolari Jupyter Notebooks (Python e R), permette ai data scientist di analizzare i dati aggregati in un ambiente sicuro e completamente integrato (Figura 4).

Nella fase di sviluppo del metodo e dell'algoritmo, gli utenti possono sviluppare le proprie pipeline, o modificare quelle esistenti, in un ambiente sandbox. Possono rapidamente costruire, testare e iterare i modelli di apprendimento automatico in base alle necessità. Gli utenti hanno accesso a una vasta gamma di librerie standard, come TensorFlow<sup>4</sup> o scikit-learn,<sup>5</sup> e possono facilmente introdurre le proprie librerie personalizzate. Quando gli utenti sono pronti a passare alla fase di produzione, la piattaforma ICA permette la conversione dei quaderni in strumenti. Questi strumenti saranno poi disponibili nel repository degli strumenti ICA e incorporati nelle pipeline di produzione.

Figura 3. La piattaforma ICA permette l'aggregazione dei dati, l'estrazione e l'apprendimento continuo: gli utenti possono esplorare le connessioni tra i set di dati per trovare le risposte alle loro domande.

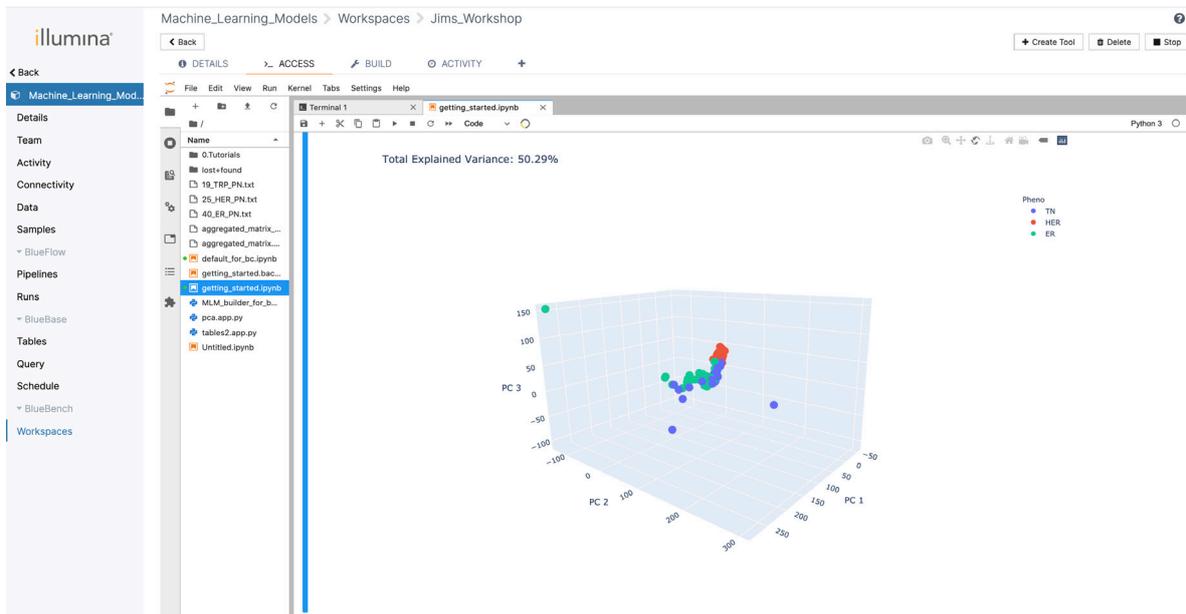


Figura 4. Analisi interattiva e visualizzazione: la piattaforma ICA supporta l'uso di Jupyter Notebooks per l'esplorazione visiva dei dati multidimensionali.

## Sicurezza al primo posto con supporto di conformità

La sicurezza è di fondamentale importanza quando si lavora con i dati in un ambiente basato sul cloud. La piattaforma ICA impiega diverse misure fisiche, elettroniche e amministrative per soddisfare anche i requisiti di sicurezza dei dati più stringenti.

- I dati caricati dagli strumenti di sequenziamento sono crittografati utilizzando lo standard AES 256 e sono protetti dal protocollo TLS (Transfer Layer Security).
- I dati della piattaforma ICA si trovano in hosting su Amazon Web Services (AWS) e devono sottostare alle migliori pratiche AWS Well-Architected, conformi a numerosi standard di sicurezza accettati nel settore.<sup>6</sup>
- Il servizio di autenticazione è supportato da SAML 2.0 per la gestione di utenti e password istituzionali (opzionale).
- I report di audit garantiscono la tracciabilità della provenienza dei dati.

La piattaforma ICA supporta anche i clienti che operano in ambienti regolamentati e devono rispettare requisiti rigorosi.

- Leggi attuali sulla protezione dei dati come il Regolamento generale sulla protezione dei dati (GDPR)<sup>7</sup> e l'Health Insurance Portability and Accountability Act (HIPAA)<sup>8</sup>
- ISO (International Organization for Standardization) 13485 sistemi di gestione della qualità<sup>9</sup> e ISO 27001 sistemi di gestione della sicurezza delle informazioni<sup>10</sup>
- Residenza dei dati garantita per assicurare la soddisfazione dei requisiti normativi e di conformità locali

## Opzioni flessibili

Per assicurare flessibilità di acquisto, la piattaforma ICA è disponibile con un abbonamento annuale. Per la fatturazione si utilizzano gli iCredit in base all'uso che si fa dello strumento e dell'archiviazione.<sup>11</sup> Gli iCredit possono essere pre-acquistati o fatturati mensilmente.

## Studi multiomici scalabili

Man mano che la generazione di dati NGS diventa più veloce e meno costosa, le piattaforme di dati avanzate che consentono ai ricercatori di passare dalle letture ai report in modo semplice e scalabile acquistano sempre più importanza. Con potenti soluzioni che supportano la collaborazione internazionale centralizzando l'accesso ai dati distribuiti, pipeline pronte all'uso e personalizzabili, accesso a strumenti di data science e un ambiente sicuro in conformità alle normative mondiali, la piattaforma ICA consente agli utenti di realizzare il pieno potenziale dei dati multiomici.

## Maggiori informazioni

Per ulteriori dettagli: [www.illumina.com/ConnectedAnalytics](http://www.illumina.com/ConnectedAnalytics)

## Bibliografia

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

## Informazioni per gli ordini

Prodotto	N. di catalogo
ICA Enterprise	20038994
ICA Data Science	20044877
Illumina Analytics - 1 iCredit	20042038
Illumina Analytics - 1.000 iCredits	20042039
Illumina Analytics - 5.000 iCredit	20042040
Illumina Analytics - 50.000 iCredit	20042041
Illumina Analytics - 100.000 iCredit	20042042
Fatturazione in base al consumo	20012931

a. Non include le funzioni data science.  
b. Fornisce accesso ai Notebook (Jupyter, R) e al framework AI/ML.  
c. L'importo da non superare è rappresentato dall'importo sul preventivo.  
Le fatture dei clienti saranno emesse mensilmente in base al consumo di calcolo, archiviazione e app di terze parti, fino all'importo associato al catalogo n. 20012931.

illumina®

+1.800.809.4566 numero verde (USA) | +1.858.202.4566 telefono  
techsupport@illumina.com | www.illumina.com

© 2020 Illumina, Inc. Tutti i diritti riservati. Tutti i marchi di fabbrica sono di proprietà di Illumina, Inc. o dei rispettivi proprietari.

Per specifiche informazioni sui marchi, consultare  
[www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).

Pubbl. n. 986-2020-009-A ITA. QB11606.