

Analisi secondaria di DRAGEN™

Identificazione di varianti accurata,
esaustiva ed efficiente con i dati
del sequenziamento di nuova
generazione

illumina®

Introduzione

Per progredire nel campo della ricerca e della medicina è fondamentale sfruttare al massimo tutte le potenzialità del genoma, grazie al sequenziamento di nuova generazione (NGS, Next-Generation Sequencing). Per utilizzare al meglio le informazioni genetiche raccolte durante il sequenziamento NGS, i ricercatori necessitano di strumenti di analisi dei dati in grado di tradurre, in modo accurato ed efficiente, le informazioni di sequenziamento non elaborate in risultati pertinenti. Inoltre, per mettere a frutto i vantaggi della tecnologia NGS, le organizzazioni necessitano di soluzioni facili da usare, adatte a ogni tipo di utente e in grado di ridurre le barriere finanziarie e tecniche legate all'adozione della tecnologia.

L'analisi secondaria di DRAGEN (Dynamic Read Analysis for GENomics) di Illumina è stata sviluppata per andare incontro alle sfide importanti poste dall'analisi dei dati NGS per una varietà di applicazioni, inclusi gli studi sul genoma, sull'esoma, sul trascrittoma e sul metiloma. La piattaforma DRAGEN è una suite di software di analisi secondaria che elabora i dati NGS e consente l'analisi terziaria per ottenere una panoramica approfondita. Gli strumenti disponibili forniscono una soluzione altamente accurata, completa ed efficiente che consente ai laboratori di qualsiasi dimensione e disciplina di ottenere di più dai propri dati genomici.

Risultati accurati

L'analisi secondaria di DRAGEN genera risultati straordinariamente accurati. Nella Precision FDA Truth Challenge V2 (PrecisionFDA V2) del 2020, DRAGEN v3.7, per i dati del sequenziamento Illumina, è stata premiata come tecnologia più accurata per tutte le regioni comparative e per tutte le regioni difficili da mappare.^{1,2} Le innovazioni introdotte in Graph Genomes e in Illumina Machine Learning (ML) con il software DRAGEN 4.0 hanno dimostrato una straordinaria accuratezza dei dati su tutte le tecnologie di sequenziamento, ottenendo il punteggio F1 del 99,83% (misurazione combinata di precisione e identificazione ripetuta) in tutte le regioni comparative (Figura 1).^{1,2} DRAGEN 4.0 + Graph (ML abilitato per impostazione predefinita) ha inoltre ottenuto il punteggio F1 più alto nell'identificazione più accurata nelle regioni del complesso maggiore di istocompatibilità (MHC, Major Histocompatibility Complex) rispetto a tutte le altre tecnologie presentate per il PrecisionFDA V2.

Analisi esaustiva

L'analisi secondaria di DRAGEN soddisfa le esigenze dei laboratori che eseguono un'ampia gamma di applicazioni NGS e offre, in una singola piattaforma, una copertura esaustiva per vari tipi di esperimenti. Le pipeline di DRAGEN supportano diversi tipi di esperimenti,

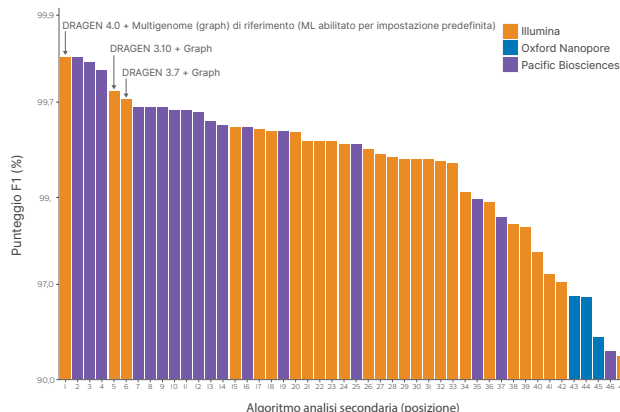


Figura 1: Accuratezza di DRAGEN 4.0 + Graph (ML attivato come impostazione predefinita) rispetto ad altri produttori nel PrecisionFDA Truth Challenge v2 per i set di dati di tutte le regioni comparative: il DRAGEN 4.0 + Graph (ML attivato come impostazione predefinita) dimostra straordinaria accuratezza assieme a Google DeepVariant sui dati di sequenziamento Pacific Biosciences. DRAGEN 3.10 + Graph mostra un potenziamento rispetto a DRAGEN 3.7 + Graph grazie ai miglioramenti nei grafici e nei riferimenti/nella gestione di contig alternati. L'asse Y contiene il punteggio F1 (%) che è un calcolo dei risultati veri positivi e veri negativi sotto forma di proporzione dei risultati finali.^{3,4}

inclusi sequenziamento dell'intero genoma (WGS, Whole-Genome Sequencing), pannelli di arricchimento, RNA-Seq di singole cellule, ATAC-Seq di singole cellule, multiomica di singole cellule, RNA-Seq collettivo e analisi di metilazione (Tabella 1). Per replicare parzialmente l'ampia gamma di funzionalità del software DRAGEN sarebbero necessari 30 strumenti open-source.^{3,4}

Il software DRAGEN offre un'ampia copertura genomica grazie alla suite di Variant Caller inclusa, che comprende: espansione di ripetizione, variazione strutturale (SV, Structural Variation), variazione del numero di copie (CNV, Copy Number Variation), ExpansionHunter e Caller mirati come *SMN*, *GBA*, *CYP2B6*, *CYP2D6* e *HLA*. Inoltre, il riferimento di DRAGEN Multigenome (Graph) estende in modo efficace le letture di Illumina ed è in grado di raggiungere le regioni a bassa complessità, risolvendo aree del genoma difficili da valutare a causa di sequenze di ripetizioni. Questo consente di migliorare la copertura di geni potenzialmente rilevanti dal punto di vista medico e di identificare singoli nucleotidi, numero di copie e varianti strutturali in regioni difficili da mappare.

Tabella 1: L'analisi secondaria di DRAGEN supporta un ampio gruppo di applicazioni di analisi secondaria

Applicazione	DRAGEN Server in laboratorio	Integrazione della serie NovaSeq X	Integrazione di NextSeq 1000 System, NextSeq 2000 System	BaseSpace Sequence Hub	Illumina Connected Analytics	
					Preconfigurato	Personalizzato
Conversione BCL	✓	✓	✓	✓	✓	✓
Compressione DRAGEN ORA	✓	✓	✓			✓
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓	✓
Genoma intero	Germline + somatico	Solo Germline	Solo Germline	Germline + somatico	Germline + somatico	Germline + somatico
Enrichment (incluso l'esoma)	Germline + somatico	Germline + somatico	Germline + somatico	Germline + somatico	Germline + somatico	Germline + somatico
DNA Amplicon	✓		✓	✓	✓	✓
RNA	✓	✓	✓	✓	✓	✓
Single-Cell RNA	✓		✓	✓	✓	✓
Espressione differenziale		✓	✓	✓		
NanoString GeoMx NGS				✓		
RNA Amplicon	✓			✓	Disponibile a breve	Disponibile a breve
Metilazione	✓			✓	✓	✓
Metagenomica				✓		
Rilevamento RNA patogeno				✓		
COVID lineage	✓			✓	Disponibile a breve	
TruSight Oncology 500	✓				✓	
ScATAC-Seq	✓			✓	✓	✓
Imputazione	✓			✓	✓	✓
PGx Star Allele Caller	✓			✓	✓	✓
Illumina Complete Long Reads				✓		

Analisi efficiente

Il software DRAGEN è stato progettato proprio per consentire ai laboratori di velocizzare l'analisi dei dati e di avere a disposizione opzioni di file necessari per sfruttare al massimo i set di dati NGS. L'accelerazione hardware dell'analisi secondaria di DRAGEN e l'architettura degli array di porte logiche programmabili sul campo (FPGA, Field-Programmable Gate Array) consentono di ottenere una rapida elaborazione. L'efficienza degli algoritmi di analisi di DRAGEN ha consentito di ottenere due record mondiali di velocità per l'analisi dei dati genomici.^{5,6} Nelle applicazioni pratiche, l'analisi secondaria di DRAGEN può elaborare in loco i dati NGS per l'equivalente di un genoma intero di 34× di copertura in circa 30 minuti, rispetto a > 15 ore impiegate da un sistema tradizionale basato su CPU.⁷

Per gestire la richiesta di archiviazione di grossi file di dati NGS, la tecnologia DRAGEN Original Read Archive (ORA) fornisce 5× di compressione dei dati dei file FASTQ senza perdita di dati. La compressione senza perdita di dati di DRAGEN ORA mantiene i dettagli dei file FASTQ ed è molto rapida, infatti richiede circa 8 minuti per la compressione di file FASTQ da 50 GB-70 GB. L'analisi secondaria di DRAGEN offre un gruppo di pipeline versatili che sono inoltre in grado di gestire file di dati di input e di creare file di output in diverse fasi della pipeline (Figura 2).

FPGA e accelerazione hardware

La tecnologia FPGA è altamente configurabile e consente l'implementazione accelerata da hardware ultra efficiente di algoritmi di analisi genomica, come conversione di file di identificazione delle basi (BCL), mappatura, allineamento, ordinamento, marcatura dei duplicati e identificazione di varianti dell'aplotipo. La natura flessibile della tecnologia FPGA consente a Illumina di sviluppare un'ampia suite di pipeline di applicazioni DRAGEN, con frequenti aggiornamenti e aggiunte per fornire la migliore accuratezza, completezza ed efficienza possibili.

Riferimenti del cliente

Il DRAGEN Reference Builder consente agli utenti di generare riferimenti umani, non umani o non standard. Questi riferimenti sono denominati tabelle hash. I riferimenti creati possono essere utilizzati come input per tutte le applicazioni di DRAGEN che supportano i file di riferimento del cliente. L'applicazione DRAGEN Reference Builder sul BaseSpace™ Sequence Hub richiede un file FASTA. La maggior parte delle pipeline di DRAGEN include il supporto integrato per hg19, hg38 (con o senza HLA*), GRCh37 e hs37d5. Con il DRAGEN Graph Toolkit, gli utenti possono ampliare le funzionalità relative alle visualizzazioni grafiche dei riferimenti anche per diverse rappresentazioni grafiche dei riferimenti umani.

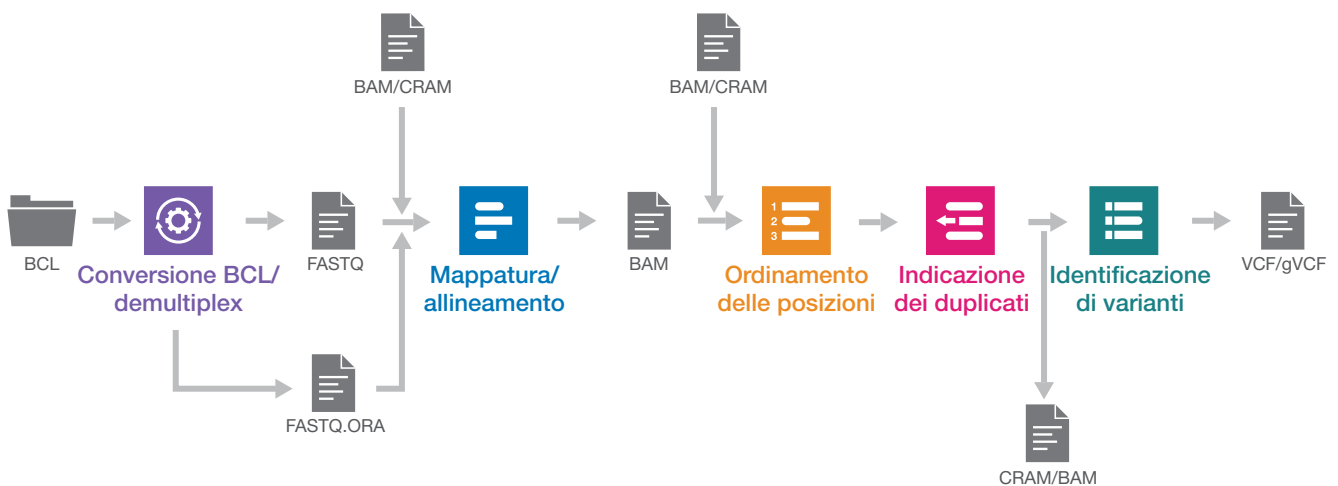


Figura 2: Flessibilità delle pipeline DRAGEN: ogni pipeline DRAGEN contiene un set specifico di passaggi che supporta un'analisi accurata ed efficiente. La pipeline DRAGEN fornisce la flessibilità di utilizzare diversi file di input e generare diversi tipi di output, in questo modo gli utenti possono personalizzare la propria esperienza e generare il formato di file desiderato.

Scalabilità

L'analisi secondaria di DRAGEN consente ai laboratori di scalare il lavoro in base alle esigenze mantenendo al contempo costi e tempi di elaborazione bassi. Il software DRAGEN può contribuire all'ampliamento delle funzionalità di ricerca in diversi modi:

- 1. Tenere il passo con la serie NovaSeq™ X:** DRAGEN integrato può, con una cella a flusso in una singola corsa, eseguire contemporaneamente fino a quattro applicazioni.
- 2. Massima capacità:** nei periodi in cui aumenta il carico di lavoro a causa di elevati volumi di campioni, i laboratori possono aumentare la capacità mediante le opzioni di accesso disponibili del software DRAGEN parallelo (Figura 3).
- 3. Ampliamento delle operazioni:** una singola istanza DRAGEN consente di eseguire tutte le pipeline DRAGEN e di elaborare tutti i tipi di campione supportati. L'accuratezza, la completezza e l'efficienza del software DRAGEN consentono agli utenti di scalare le operazioni senza compromissione dei tempi di elaborazione o della qualità dei risultati.
- 4. Dagli esomi ai genomi:** il passaggio dal sequenziamento dell'intero esoma (WES, Whole-Exome Sequencing) al sequenziamento dell'intero genoma (WGS, Whole-Genome Sequencing) comporta un significativo aumento dei dati generati. Il software DRAGEN consente ai clienti di eseguire le analisi dagli esomi ai genomi senza grossi investimenti in un'ulteriore infrastruttura hardware o soluzioni basate sul cloud.
- 5. Set di dati molto grandi:** l'analisi secondaria di DRAGEN offre un flusso di lavoro semplificato per l'analisi di coorti su larga scala; quando vengono utilizzate più pipeline assieme è possibile identificare varianti piccole e grandi con elevata accuratezza da un campionamento delle coorti. Il software DRAGEN consente l'aggregazione e la genotipizzazione da migliaia a milioni di file nel formato di identificazione di varianti del genoma (gVCF, genome Variant Call Format) e aggrega nuovi batch senza dover rielaborare i batch esistenti. La pipeline DRAGEN Joint Genotyping identifica congiuntamente le varianti su più genomi e scala a grandi coorti con una rapida analisi e senza compromettere l'accuratezza.⁸ L'analisi secondaria di DRAGEN sui dati di [1000 Genomes Project](#), ad esempio, consente l'identificazione di varianti accurate e su larga scala di diversi campioni e l'identificazione di regioni in cui i dati di copertura non sono uniformi o deviano dalle ipotesi.

Accessibilità su più piattaforme

Alla suite delle pipeline DRAGEN si accede tramite le soluzioni disponibili in laboratorio, integrate sullo strumento o sul cloud, che consentono ai laboratori di selezionare la soluzione che meglio soddisfa le proprie esigenze (Figura 3).

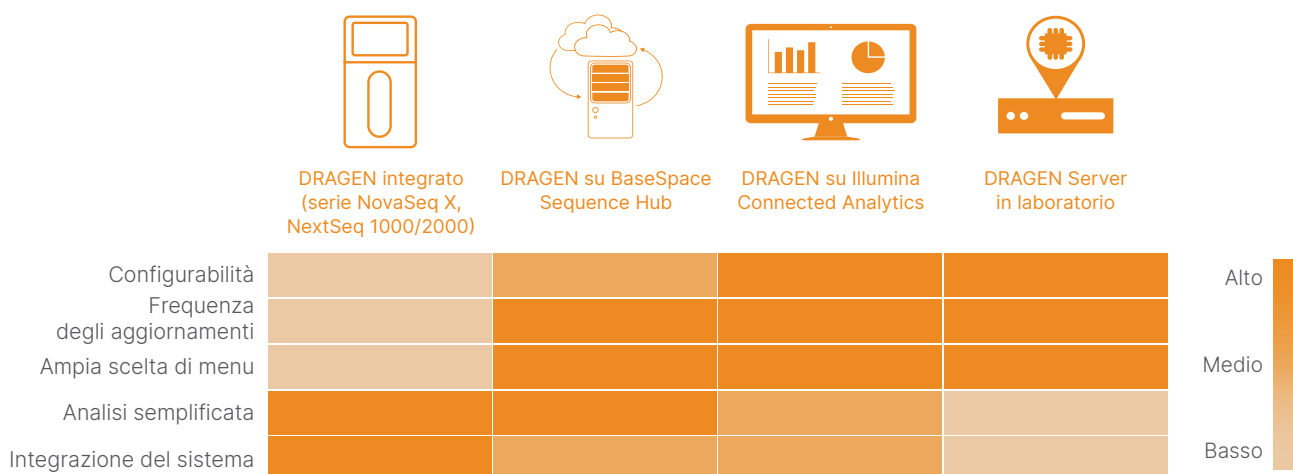


Figura 3: Opzioni di accesso della pipeline DRAGEN con caratteristiche progettate per soddisfare le esigenze di ogni laboratorio per l'analisi dei dati NGS.

DRAGEN Server in laboratorio

DRAGEN in laboratorio si affida a una soluzione di archiviazione locale per raccogliere e archiviare i dati NGS. Dopo che i dati del sequenziamento non elaborati sono stati trasferiti dallo strumento di sequenziamento all'archivio locale mediante una connessione di rete locale, i dati vengono trasferiti al DRAGEN Server per eseguire il flusso di lavoro selezionato. Dopo l'analisi, il software scrive i file di output generati nella posizione di archiviazione locale. DRAGEN Server in laboratorio:

- Supporta diversi livelli di interfaccia con linea di comando.
- Sostituisce fino a 30 istanze di calcolo tradizionali.
- Elabora i dati NGS per un intero genoma umano a una copertura di 34x in circa 30 minuti.

DRAGEN incluso nella serie NovaSeq™ X

La serie NovaSeq X include il più potente software DRAGEN progettato fino ad ora e offre un'analisi secondaria accurata, automatizzata ed esaustiva. La suite di software DRAGEN integrata fornisce identificazione di varianti e compressione ORA per le applicazioni NGS (Tabella 1) inclusi BCL Convert, Germline, RNA ed Enrichment. Dragen integrato:

- Esegue in parallelo diverse pipeline di analisi secondaria.
- Con una cella a flusso in una singola corsa esegue contemporaneamente fino a quattro applicazioni.
- Aumenta fino a 5x la compressione di dati senza perdita di dati e l'analisi per le applicazioni supportate.
- Consente di risparmiare sull'analisi che nell'arco di più di cinque anni può superare il costo dell'acquisto del NovaSeq X System.

DRAGEN integrato in NextSeq™ 1000 System e NextSeq 2000 System

Il NextSeq 1000 System e il NextSeq 2000 System includono il software DRAGEN integrato per un'analisi secondaria accurata e rapida. La suite di software DRAGEN integrata offre una serie selezionata di pipeline per una gamma di applicazioni NGS comuni (Tabella 1) con un'interfaccia utente di facile utilizzo che consente agli utenti esperti e non esperti di eseguire le analisi necessarie e generare risultati in modo rapido.

Dragen integrato:

- Consente di accedere a pipeline informatiche DRAGEN selezionate.
- Consente agli utenti di generare risultati in appena due ore.
- Utilizza algoritmi della pipeline intuitivi per ridurre la necessità di affidarsi a esperti informatici esterni.

BaseSpace Sequence Hub

La suite DRAGEN basata sul cloud è disponibile su BaseSpace Sequence Hub e unisce analisi accurata ed efficiente con un ecosistema sicuro e scalabilità versatile. Il software DRAGEN su BaseSpace Sequence Hub semplifica l'analisi secondaria per i laboratori di tutte le dimensioni e discipline. BaseSpace Sequence Hub è un'estensione diretta dei vostri strumenti di Illumina. I dati codificati vengono trasferiti dallo strumento a BaseSpace Sequence Hub e un set di applicazioni selezionate semplifica l'analisi e la gestione dei dati. BaseSpace Sequence Hub, alimentato da Amazon Web Services (AWS):

- Offre una soluzione semplice e di facile utilizzo per l'analisi DRAGEN.
- Utilizza un'interfaccia grafica utente intuitiva che consente a utenti esperti e non esperti di lavorare in modo efficiente.
- Fornisce accesso a fonti di calcolo efficaci senza spese di capitale per l'acquisto di ulteriore infrastruttura.

Illumina Connected Analytics

L'analisi secondaria di DRAGEN sull'Illumina Connected Analytics è una piattaforma bioinformatica completa e basata sul cloud che consente ai ricercatori di gestire, analizzare e interpretare grandi volumi di dati multiomici in un ambiente sicuro, scalabile e flessibile. Illumina Connected Analytics:

- Consente di accedere al software DRAGEN completo, disponibile in pipeline preconfezionate o singoli strumenti per le pipeline del cliente.
- Supporta flussi di lavoro altamente automatizzati e soluzioni personalizzate per studi ottimizzati e a elevata processività.
- Offre un ambiente altamente sicuro con residenza di dati garantita, accesso singolo (SSO, Single Sign-On), registri di audit e controllo dell'accesso che supporta la conformità a Health Insurance Portability and Accountability Act (HIPAA) e ai principi del regolamento generale sulla protezione dei dati (RGPD) dell'Unione Europea.

Riepilogo

L'analisi secondaria di DRAGEN è una suite efficace di strumenti software che fornisce un'analisi accurata, esaustiva ed efficiente dei dati NGS. I laboratori possono selezionare tra le diverse opzioni di software DRAGEN e scegliere la soluzione che meglio si adatta al tipo e alla scala dei propri progetti. La tecnologia NGS continua a progredire e gli aggiornamenti tempestivi apportati all'analisi secondaria di DRAGEN consentono di ottenere le migliori prestazioni possibili dalle pipeline attuali, mentre nuove pipeline continuano a essere aggiunte man mano che le applicazioni diventano disponibili.

Maggiori informazioni

[Analisi secondaria di DRAGEN](#)

[Pagina di supporto dell'analisi secondaria di DRAGEN](#)

[Contatti](#)

Bibliografia

1. Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. precision.fda.gov/challenges/10. Consultato il 14 marzo 2022.
2. Illumina. DRAGEN Sets New Standard for Data Accuracy in PrecisionFDA Benchmark Data. Optimizing Variant Calling Performance with Illumina Machine Learning and DRAGEN Graph. illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html. Consultato il 14 marzo 2022.
3. Illumina. DRAGEN Wins at PrecisionFDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html. Consultato il 14 marzo 2022.
4. Dati interni in archivio. Illumina, Inc. 2022.
5. BioIT World. Children's Hospital Of Philadelphia, Edico Set World Record For Secondary Analysis Speed. bio-itworld.com/news/2017/10/23/children-s-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed. Consultato il 14 marzo 2022.
6. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.sandiegouniontribune.com/95899028-132.html>. Pubblicato il 12 febbraio 2018. Consultato il 14 marzo 2022.
7. Miller NA, Farrow EG, Gibson M, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* 2015;7:100. doi: 10.1186/s13073-015-0221-8.
8. Illumina. Accurate and Efficient Calling of Small and Large Variants from PopGen Datasets Using the DRAGEN Bio-IT Platform. www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. Consultato il 14 marzo 2022.

illumina®

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

© 2023 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-00680-ITA v5.0