

Piattaforma DRAGEN[™] Bio-IT integrata sui sistemi NextSeq[™] 1000 e NextSeq 2000

Soluzione accurata,
efficace e integrata
per l'analisi dei dati NGS

- Flusso di lavoro ottimizzato per analisi secondaria altamente accurata con interfaccia grafica di facile utilizzo
- Generazione di identificazioni di varianti di alta qualità integrata direttamente sui sistemi NextSeq 1000 e NextSeq 2000
- Riduce la necessità di ulteriore infrastruttura bioinformatica e richiede minori touchpoint rispetto ad altri software DRAGEN

illumina[®]

Introduzione

L'obiettivo condiviso di molti laboratori in tutto il mondo è quello di svelare la potenza del genoma. I recenti progressi nella tecnologia hanno migliorato l'accesso a piattaforme e metodi di sequenziamento di nuova generazione (NGS, Next-Generation Sequencing) che possono contribuire a svelare i segreti del genoma. I sistemi NextSeq 1000 e NextSeq 2000 supportano applicazioni con bassa e media processività, inclusi il sequenziamento di esomi, trascrittomi e genomi piccoli, arricchimento target, sequenziamento di singole cellule e metagenomica shotgun.

Uno dei principali vantaggi dei sistemi NextSeq 1000 e NextSeq 2000 è la possibilità di includere un'implementazione integrata della piattaforma DRAGEN (Dynamic Read Analysis for GENomics) Bio-IT. Il software DRAGEN integrato fornisce ai laboratori l'analisi secondaria dei dati NGS straordinariamente completa, accurata e veloce, riducendo la necessità di ulteriore infrastruttura di calcolo e risorse bioinformatiche. Con il software DRAGEN integrato, l'analisi secondaria può essere configurata durante l'impostazione della corsa di sequenziamento; ciò consente di utilizzare un flusso di lavoro ottimizzato con un numero inferiore di touchpoint e un tempo di elaborazione più veloce. Per i laboratori che sono alla ricerca di un'elevata configurabilità, Illumina offre ulteriori opzioni per l'implementazione del software DRAGEN (Figura 1).

Il software DRAGEN integrato è perfetto per i sistemi versatili e scalabili NextSeq 1000 e NextSeq 2000. Grazie all'utilizzo di diversi algoritmi all'avanguardia, il software DRAGEN integrato consente agli investigatori di superare le difficoltà poste dall'analisi dei dati e di generare velocemente risultati accurati per un'ampia gamma di applicazioni (Tabella 1).

Efficienza

Il software DRAGEN integrato sui sistemi NextSeq 1000 e NextSeq 2000 consente ai laboratori di integrare in modo ottimale le funzioni di sequenziamento e analisi durante l'impostazione della corsa. Il software DRAGEN integrato consente ai laboratori di ottenere file FASTQ e file nel formato di identificazione di varianti (VCF) da usare nelle applicazioni a valle direttamente dallo strumento mentre la corsa viene completata, risparmiando tempo nel trasferimento dei dati e nell'analisi. Gli algoritmi della pipeline DRAGEN integrata consente a utenti novizi ed esperti di completare le funzioni di analisi comuni e di ridurre il bisogno di affidarsi a esperti informatici esterni. Grossi file di dati possono essere gestiti con l'archiviazione dei file DRAGEN Original Read Archive (ORA), che permette di ottenere una media di compressione di file FASTQ di 4x senza perdita di dati e di utilizzare l'80% in meno di energia per un luogo di lavoro più sostenibile.



Figura 1: Opzioni di implementazione della pipeline DRAGEN Bio-IT per soddisfare le esigenze di ogni laboratorio per l'analisi dei dati NGS.

Tabella 1: Applicazioni del software DRAGEN integrato sui sistemi NextSeq 1000 e NextSeq 2000

Applicazione ^a	Descrizione
Conversione BCL	Converte file BCL generati dai sistemi di sequenziamento Illumina in file FASTQ.
Compressione DRAGEN ORA	Genera compressione di file FASTQ basati su riferimento e senza perdita di dati.
DRAGEN FASTQ + MultiQC	Esegue le metriche FastQC con accelerazione hardware senza perdere ulteriore tempo.
Genoma intero	Esegue la mappatura del genoma umano, l'allineamento e l'identificazione di varianti piccole. Solo linea germinale.
Arricchimento (incluso l'esoma)	Esegue l'identificazione di varianti piccole in campioni della linea germinale oppure l'identificazione di varianti a bassa frequenza in campioni somatici.
DNA Amplicon	Analizza la variazione genetica in determinate regioni genomiche. Utilizza la pipeline DRAGEN DNA con una fase in più per sottoporre i primer a soft-clipping e riscrivere gli allineamenti. Ciò assicura che le sequenze dei primer non contribuiscono alle identificazioni delle varianti.
RNA	Offre uno strumento di allineamento RNA-Seq (attento alle varianti di splicing) con filtraggio facoltativo di rRNA durante l'allineamento, riducendo la durata e la dimensione dei file.
Single-Cell RNA	Elabora un'ampia gamma di set di dati RNA-Seq a singola cellula ottenuti dalle letture su matrici di espressione genica di cellula per gene.
Espressione differenziale	Esegue l'algoritmo DESeq2 sui dati di quantificazione dell'RNA generati dalla pipeline DRAGEN RNA. Fornisce geni e trascritti espressi differenzialmente tra due gruppi di campioni.
NanoString GeoMx NGS	Ottimizza l'analisi per i clienti che utilizzano sia GeoMx NGS sia gli strumenti Illumina per flussi di lavoro di genomica spaziale.
Metilazione	Gestisce i dati di metilazione di bisolfiti e di TET-Assisted Pyridine Sequencing (TAPS).

a. Ulteriori pipeline di applicazioni sono disponibili in BaseSpace Sequence Hub, Illumina Connected Analytics, DRAGEN Multi-Cloud e DRAGEN server in laboratorio.

Accuratezza

Le pipeline di analisi DRAGEN generano risultati straordinariamente accurati. Nella Precision FDA Truth Challenge V2 (PrecisionFDA V2) del 2020, DRAGEN v3.7, per i dati del sequenziamento Illumina, ha vinto per i dati più accurati per tutte le regioni comparative e per tutte le regioni difficili da mappare. Ulteriori innovazioni in Graph Genomes e in Illumina Machine Learning con DRAGEN v3.10 hanno stabilito un nuovo standard di accuratezza dei dati in tutte le tecnologie di sequenziamento, ottenendo il punteggio F1 del 99,83% in tutte le regioni comparative (un calcolo dei risultati veri positivi e veri negativi genera una proporzione dei risultati totali) (Figura 2). DRAGEN 3.10, Graph e Illumina Machine Learning hanno inoltre ricevuto il punteggio F1 più alto per l'identificazione più accurata nelle regioni del complesso maggiore di istocompatibilità (MHC, Major Histocompatibility Complex) rispetto a tutti gli altri strumenti che hanno partecipato alla PrecisionFDA V2.^{1,2}

La straordinaria efficacia della piattaforma DRAGEN Bio-IT consente di analizzare rapidamente i dati grazie all'utilizzo della tecnologia delle matrici di porte logiche programmabili sul campo (FPGA, Field-Programmable Gate Array). La scheda FPGA integrata, unica per i sistemi NextSeq 1000 e NextSeq 2000, consente l'utilizzo di pipeline informatiche per l'analisi secondaria accelerata da hardware per diverse applicazioni. Queste pipeline DRAGEN sono in continuo miglioramento e le nuove pipeline offrono le migliori funzionalità, accuratezza e velocità. La piattaforma è progettata proprio per le esigenze di analisi degli investigatori e consente di sprecare meno tempo e sforzi nella gestione di pipeline a livello di produzione e concentrarsi invece sui risultati.

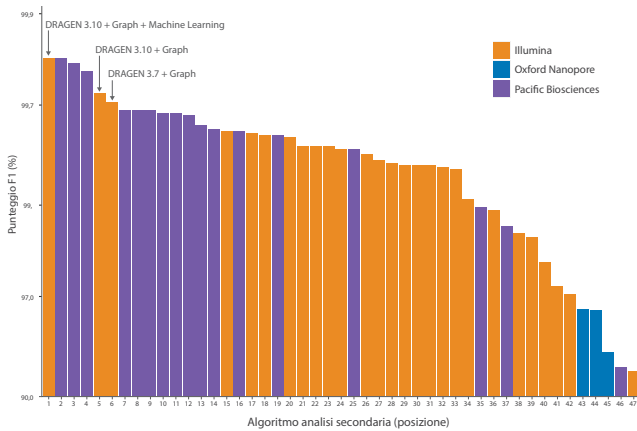


Figura 2: Accuratezza nell'identificazione di varianti piccole di DRAGEN 3.10, Graph e Machine Learning rispetto ad altri produttori nel PrecisionFDA Truth Challenge v2 per i set di dati di tutte le regioni comparative: il metodo di analisi DRAGEN 3.10 + Graph + Machine Learning ha ottenuto il primo posto assieme agli strumenti meglio performanti di Pacific Biosciences. DRAGEN 3.10 + Graph mostra un miglioramento rispetto a DRAGEN 3.7 + Graph grazie ai miglioramenti nei grafici e nei riferimenti/gestione di contig alternati.

Ampia selezione di applicazioni

Il software DRAGEN integrato sui sistemi NextSeq 1000 e NextSeq 2000 dispone di pipeline di analisi versatili e genera diversi file di output in diverse fasi del flusso di lavoro (Figura 3). La selezione di pipeline di analisi inclusa supporta un'ampia gamma di tipi di analisi, incluse in via esemplificativa l'analisi di singola cellula, esoma e RNA (Tabella 1).

Pipeline DRAGEN Single-Cell RNA integrate

Le pipeline DRAGEN Single-Cell RNA sono soluzioni rapide e scalabili in grado di elaborare un'ampia gamma di set di dati di RNA-Seq a singola cellula con formati di output di facile utilizzo, inclusi grafici di cluster per tipo di cellula basati sulle attività di espressione dell'RNA (Figura 4). Inoltre, dispone di utili estensioni per l'elaborazione di set di dati in multiplex che consistono di diversi campioni (ad es. utilizzo di demultiplex del genotipo o hashing della cellula) e il conteggio dell'espressione delle proteine della superficie cellulare. Supporta inoltre una gamma di tipi di preparazione delle librerie di input per la compatibilità con strumenti di analisi a valle.

Il controllo qualità integrato delle librerie di espressione di singola cellula e la pipeline di analisi di singola cellula vengono eseguiti in tempi più brevi rispetto alle soluzioni basate sul cloud e senza sacrificare l'accuratezza. La pipeline per singola cellula consente di partire dall'impostazione della corsa fino all'espressione quantificata per cellula sugli strumenti NextSeq 1000 e NextSeq 2000 da un singolo touchpoint, riducendo la necessità di ulteriori risorse informatiche.

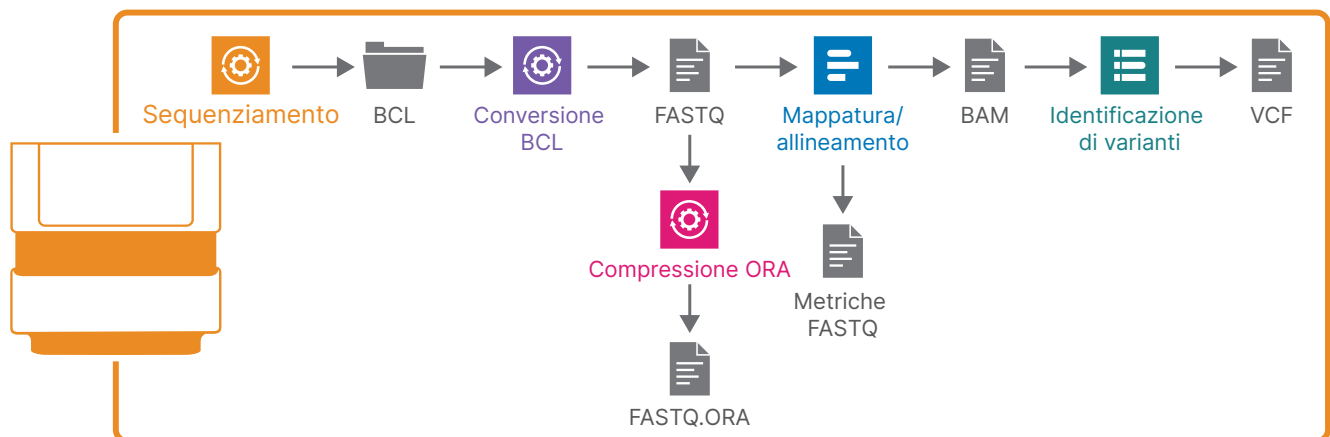


Figura 3: Il software DRAGEN integrato sui sistemi NextSeq 1000 e NextSeq 2000 fornisce un flusso di lavoro ottimizzato per l'elaborazione dei dati e l'analisi secondaria: la soluzione software DRAGEN integrata fornisce uno strumento molto efficace per generare risultati nei formati richiesti dai laboratori per le applicazioni a valle.

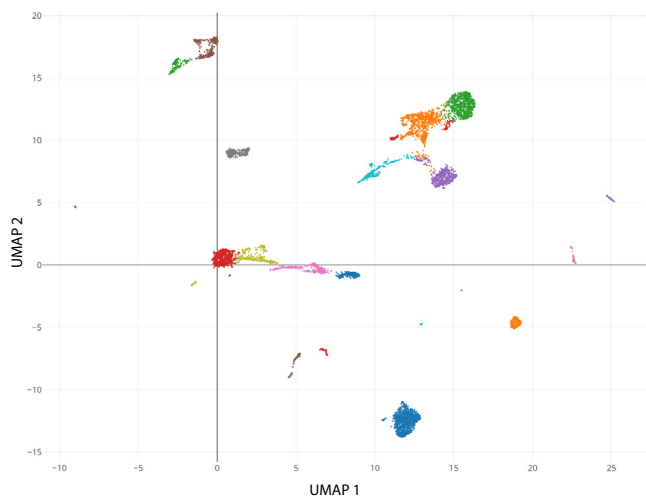


Figura 4: Esempio di grafico di cluster per tipo di cellula ottenuto dalla pipeline DRAGEN Single Cell RNA: l'analisi UMAP (Uniform Manifold Approximation and Projection) consente la visualizzazione di singole cellule per tipo o per stato funzionale in base a profili di espressione multidimensionali complessi.⁴

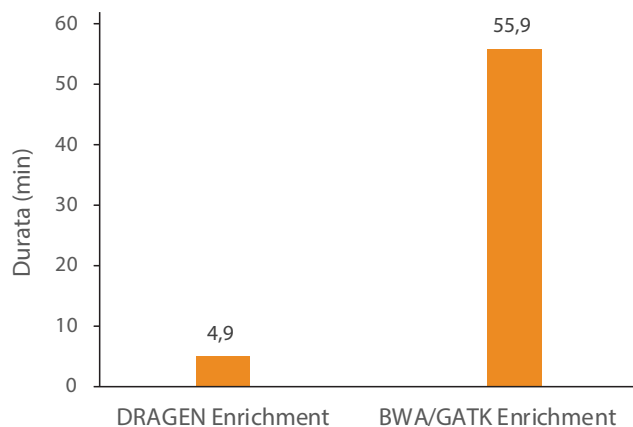


Figura 5: Durata media per campione della pipeline DRAGEN Enrichment e dell'applicazione BWA Enrichment su BaseSpace™ Sequence Hub: l'analisi dell'arricchimento della linea germinale di 24 replicati NGS mostra che la pipeline DRAGEN Enrichment offre tempi di analisi per campione più rapidi rispetto a simili analisi eseguite con BWA Enrichment seguite dall'identificazione di varianti con GATK Variant Caller.

Pipeline DRAGEN Enrichment integrata

La pipeline DRAGEN Enrichment integrata fornisce un'analisi veloce e include la calibrazione di modelli di errore avanzati per migliorare l'accuratezza dei pannelli di arricchimento e dei pannelli di esomi basati su ibridazione nonché metriche basate sul target di input. Il software può essere eseguito in due modalità: la modalità per la linea germinale consente l'identificazione di varianti piccole della linea germinale mentre la modalità somatica consente l'identificazione di varianti somatiche (bassa frequenza). L'identificazione di variazioni del numero di copie (CNV, Copy Number Variation) e l'identificazione di varianti strutturali possono essere abilitate nella modalità della linea germinale. La pipeline DRAGEN Enrichment è più veloce e più accurata rispetto all'analisi con Burrows-Wheeler Aligner (BWA) e Genome Analysis Toolkit (GATK) Variant Caller, in particolar modo nell'identificazione delle indel (Tabella 2, Figura 5). Una cella a flusso sui sistemi NextSeq 1000 e NextSeq 2000 viene di solito elaborata in meno di due ore.

Pipeline DRAGEN RNA integrata

La pipeline DRAGEN RNA integrata esegue l'analisi secondaria dei trascritti di RNA. La pipeline per l'RNA offre diverse modalità di funzionamento, inclusi l'allineamento basato solo sul riferimento e l'allineamento assistito da annotazione con il rilevamento delle fusioni geniche. Un filtro per RNA ribosomiale (rRNA) facoltativo può essere abilitato per ridurre la durata dell'analisi e la dimensione dei file. Il modulo per la fusione genica sfrutta lo strumento di allineamento di splicing DRAGEN RNA per eseguire l'analisi delle letture di separazione su ulteriori allineamenti (chimerici) per rilevare eventuali breakpoint aggiungendo un minimo tempo di elaborazione all'analisi complessiva. Grazie alle funzioni del software DRAGEN integrato è possibile eseguire il rilevamento delle fusioni geniche, la quantificazione dei trascritti e l'identificazione di varianti.

Tabella 2: Risultati di accuratezza per l'arricchimento della pipeline DRAGEN Enrichment e BWA/GATK

	SNV			Indel		
	Precisione (%)	Identificazione ripetuta (%)	F1	Precisione (%)	Identificazione ripetuta (%)	F1
DRAGEN Enrichment	99,7	94,64	97,11	96,54	83,83	89,74
BWA/GATK Enrichment	99,82	92,77	96,17	96,71	71,32	82,1

Riepilogo

Il software DRAGEN Bio-IT integrato sui sistemi NextSeq 1000 e NextSeq 2000 offre bioinformatica di facile configurazione e un'esperienza utente semplificata in un'implementazione in laboratorio. L'analisi secondaria dei dati NGS generati sugli strumenti è più veloce rispetto alle pipeline DRAGEN su BaseSpace™ Sequence Hub. Le pipeline integrate come singola cellula, arricchimento e RNA sono più veloci e accurate rispetto agli strumenti attualmente disponibili.

Informazioni per gli ordini

Prodotto	N. di catalogo
NextSeq 1000 Sequencing System	20038898
NextSeq 2000 Sequencing System	20038897
NextSeq 1000 to NextSeq 2000 upgrade	20047256

Maggiori informazioni

Sistemi di sequenziamento NextSeq 1000 e NextSeq 2000, illumina.com/systems/sequencing-platforms/nextseq-1000-2000.html.

Piattaforma DRAGEN Bio-IT, illumina.com/products/by-type/informatics-products/dragen-bio-it-platform.html.

Supporto per DRAGEN, support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform.html.

illumina®

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

© 2022 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, consultare la pagina Web www.illumina.com/company/legal.html.
M-GL-00729 v1.0 ITA

Bibliografia

1. U.S. Food & 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. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-498. doi:10.1038/ng.806
4. McInnes, L, Healy J, and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*. Revised Sep 18, 2020. Consultato il 5 maggio 2022. doi.org/10.48550/arXiv.1802.03426