

Analisi secondaria DRAGEN™

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

illumina®

Introduzione

Per progredire nei campi della ricerca biomedica e della medicina di precisione, è fondamentale sfruttare al massimo le potenzialità del genoma attraverso il sequenziamento di nuova generazione (NGS, Next-Generation Sequencing). Per utilizzare al meglio le informazioni genetiche raccolte durante l'NGS, i ricercatori necessitano di strumenti di analisi dei dati in grado di tradurre in modo accurato ed efficiente i dati di sequenziamento non elaborati in risultati significativi. Inoltre, per mettere a frutto i vantaggi dell'NGS, le organizzazioni hanno bisogno di soluzioni semplici da usare, adatte a ogni tipo di utente e facili da adottare, senza eccessive barriere finanziarie e tecniche.

L'analisi secondaria DRAGEN (Dynamic Read Analysis for GENomics) Illumina è stata sviluppata per superare le principali difficoltà associate all'analisi dei dati di NGS in una varietà di applicazioni, inclusi, tra gli altri, gli studi sul genoma, sull'esoma, sul trascrittoma e sul metiloma. Il software di analisi secondaria DRAGEN è una suite di applicazioni che consente di ottenere informazioni approfondite attraverso l'elaborazione dei dati di NGS e l'analisi terziaria. Gli strumenti disponibili compongono 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 DRAGEN genera risultati straordinariamente accurati. Nella Precision FDA Truth Challenge V2 (PrecisionFDA V2) del 2020, l'analisi secondaria DRAGEN v3.7 è stata premiata come la più accurata in tutte le regioni di riferimento e nelle regioni difficili da mappare con i dati di sequenziamento Illumina.^{1,2} Le versioni successive continuano a definire nuovi standard di accuratezza, con progressi in aree quali apprendimento automatico (ML, Machine Learning) e tecnologia multigenomica (grafico) di DRAGEN. L'ultima versione dell'analisi secondaria DRAGEN, la v4.3, offre un'accuratezza senza precedenti nell'identificazione di varianti piccole, con un punteggio F1 del 99,89% (una misura combinata di precisione e richiamo) in tutte le regioni di riferimento (figura 1). Ciò è reso possibile dal riferimento (grafico) multigenomico di terza generazione di DRAGEN, basato su 128 campioni con 256 aplotipi da dati di riferimento del pangenoma costruiti internamente, che cattura una maggiore diversità genetica. All'accuratezza superiore contribuisce anche il nuovo identificatore integrato di varianti a mosaico, che può essere abilitato a rilevare varianti a mosaico con basse frequenze alleliche (fino al 3%).

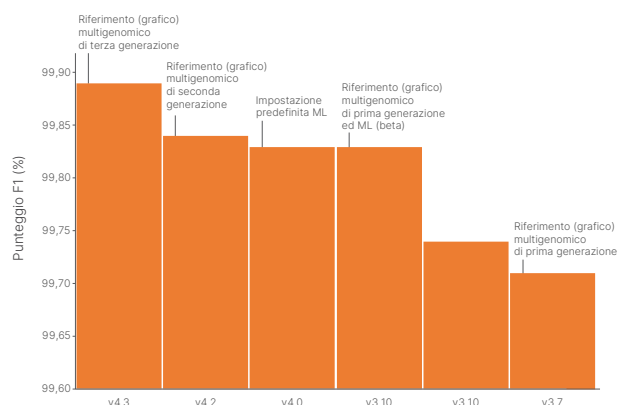


Figura 1: accuratezza dell'analisi secondaria DRAGEN.

Il punteggio F1 (%), sull'asse Y, è un calcolo dei risultati veri positivi e veri negativi sotto forma di proporzione dei risultati totali.^{3,4}

Analisi esaustiva

Con una copertura completa del genoma e un'ampia gamma di applicazioni supportate, l'analisi secondaria DRAGEN soddisfa le esigenze variegata dei laboratori che svolgono analisi di NGS. Le pipeline di DRAGEN supportano diversi tipi di esperimenti, inclusi sequenziamento dell'intero genoma (WGS, Whole-Genome Sequencing), pannelli di arricchimento, RNA-Seq di singole cellule, ATAC-Seq di singole cellule, RNA-Seq in massa 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}

Per l'analisi della linea germinale, l'analisi secondaria DRAGEN include una suite di identificatori di varianti, come ExpansionHunter, e identificatori mirati, quali *SMN*, *GBA*, *CYP2B6*, *CYP2D6* e *HLA*. DRAGEN v4.3 introduce anche l'MRJD, un nuovo identificatore specializzato che consente la copertura di geni complessi in regioni di duplicazione segmentale, come *PMS2*, *SMN1*, *SMN2*, *STRC*, *NEB*, *TTN* e *IKBKKG*. Questi strumenti permettono di analizzare un'ampia gamma di variazioni genetiche, comprese variazioni a singolo nucleotide, inserzioni e delezioni (indel), espansioni di ripetizioni e variazioni strutturali in regioni genomiche estese. Il riferimento (grafico) multigenomico di DRAGEN migliora inoltre la qualità della mappatura, rendendo più accurata l'identificazione di varianti e superando le difficoltà di valutazione, dovute alla complessità della sequenza, di alcune aree del genoma. Ciò consente di aumentare la copertura di geni potenzialmente rilevanti dal punto di vista medico e di identificare varianti a singolo nucleotide, piccole indel, variazione del numero di copie e varianti strutturali in regioni difficili da mappare.

Tabella 1: l'analisi secondaria DRAGEN supporta una gamma estesa di applicazioni di analisi secondaria^a

Applicazione	Server in laboratorio	Sistemi di sequenziamento Illumina integrati		Piattaforme cloud Illumina	
	Sever DRAGEN	NovaSeq X Series	NextSeq 1000 System, NextSeq 2000 System	BaseSpace Sequence Hub	Illumina Connected Analytics
Conversione BCL	✓	✓	✓	✓	Solo personalizzata
DRAGEN ORA Compression	✓	✓	✓		Solo personalizzata
DRAGEN FASTQ + MultiQC	✓	✓	✓	✓	✓
Genoma intero	Linea germinale + somatico	Linea germinale + somatico	Linea germinale + somatico	Linea germinale + somatico	Linea germinale + somatico
Arricchimento (incluso l'esoma)	Linea germinale + somatico	Linea germinale + somatico	Linea germinale + somatico	Linea germinale + somatico	Linea germinale + somatico
DRAGEN Amplicon	✓		Solo DNA	✓	✓
RNA	✓	✓	✓	✓	✓
RNA di singole cellule	✓		✓	✓	✓
NanoString GeoMx NGS			✓	✓	
Metilazione	✓	✓		✓	✓
Metagenomica	✓ ^b			✓	
Rilevamento RNA patogeno				✓	
COVID	COVIDSeq, COVID Lineage		COVIDSeq (solo cloud)	COVIDSeq, COVID Lineage	
Portafoglio TruSight Oncology 500	✓			✓ ^c	✓
scATAC-Seq	✓			✓	✓
Imputazione	✓			✓	✓
PGX Star Allele Caller	✓	✓	✓	✓	✓
Illumina Complete Long Reads				✓	✓
Analisi secondaria DRAGEN per RPIP e UPIP	✓			✓	Beta

a. La versione del software principale DRAGEN varia da una piattaforma all'altra; per ulteriori informazioni, consultare il rappresentante locale.

b. Applicazioni di metagenomica abilitate dal classificatore Kmer; altri strumenti presto disponibili.

c. È richiesto l'abbonamento a Illumina Connected Analytics.

Analisi efficiente

Il software DRAGEN è progettato per offrire ai laboratori la velocità di analisi dei dati di cui hanno bisogno per elaborare i set di dati di NGS con la massima efficienza. L'accelerazione hardware dell'analisi secondaria DRAGEN e l'architettura degli array di porte logiche programmabili sul campo (FPGA, Field-Programmable Gate Array) contribuiscono ad accorciare i tempi di risposta. L'efficienza degli algoritmi di analisi DRAGEN è stata alla base del raggiungimento di due record mondiali di velocità per l'analisi dei dati genomici.^{5,6} Nella pratica, l'analisi secondaria DRAGEN in laboratorio può elaborare i dati di NGS per un genoma intero, equivalenti a una copertura di 40x, in circa 35 minuti con tutti gli identificatori⁷, mentre i metodi open source comunemente utilizzati, che identificano un numero limitato di tipi di varianti, impiegano più di 8 ore.⁷

Per semplificare l'archiviazione, la gestione e la condivisione di file di dati di NGS di grandi dimensioni, la tecnologia DRAGEN Original Read Archive (ORA) offre una compressione fino a 5x e senza perdita di dati dei file FASTQ nel tradizionale formato fastq.gz.

* Secondo i dati interni di Illumina relativi al server DRAGEN v4 e in base agli standard HG001-HG007, senza nuovi identificatori specializzati, come MRJD e VNTR, disponibili in DRAGEN v4.3.

La compressione senza perdita di dati della tecnologia DRAGEN ORA preserva i dettagli dei file FASTQ ed è molto rapida, infatti richiede circa 8 minuti per la compressione di file FASTQ da 50-70 GB, oltre a supportare una gamma estesa di specie comunemente studiate. L'analisi secondaria DRAGEN è caratterizzata da pipeline versatili, che sono inoltre in grado di gestire file di dati di input e di creare file di output in diverse fasi dei processi (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 degli aplotipi. 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.

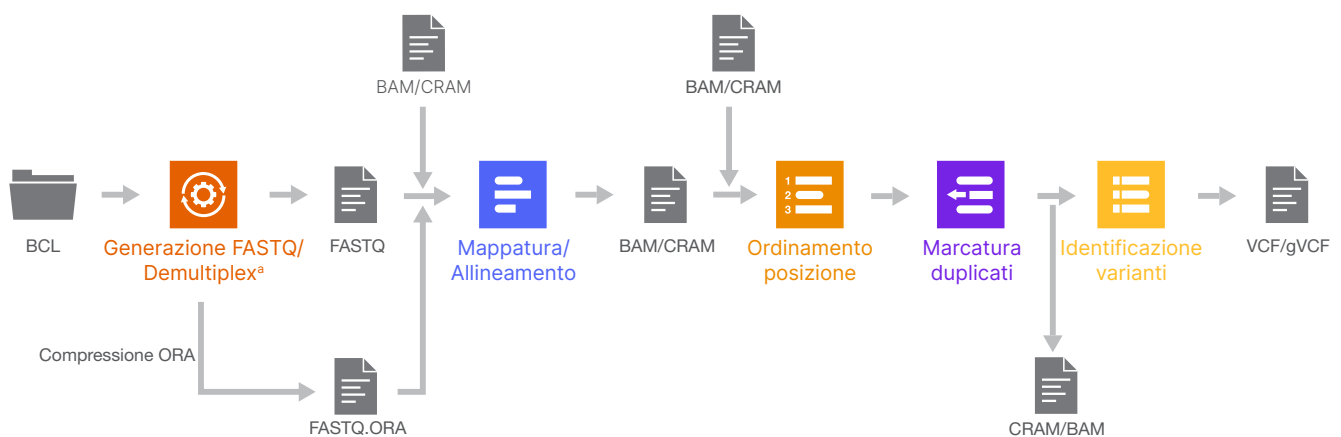


Figura 2: flessibilità delle pipeline di analisi secondaria DRAGEN. Ogni pipeline di DRAGEN contiene una serie specifica di passaggi per un'analisi accurata ed efficiente. La pipeline esemplificativa della linea germinale dell'intero genoma di DRAGEN è abbastanza flessibile da accettare diversi file di input e generare vari tipi di output; in questo modo gli utenti possono personalizzare la propria esperienza e generare il formato file desiderato.

a. La conversione BCL è disponibile anche come strumento indipendente.

Riferimenti personalizzati

L'analisi secondaria DRAGEN consente agli utenti di generare riferimenti personalizzati umani, non umani o non standard. I riferimenti creati possono essere utilizzati come input per tutte le applicazioni di DRAGEN che supportano file di riferimento personalizzati. La maggior parte delle pipeline di DRAGEN include il supporto integrato di hg19, hg38 (con o senza HLA), GRCh37, CHM13v2 e hs37d5. Il software DRAGEN consente agli utenti di estendere le capacità di creazione di riferimenti grafici multigenomici standard per popolazioni diversificate e specifiche.

Scalabilità

L'analisi secondaria DRAGEN consente ai laboratori di ampliare le procedure operative in base alle esigenze, riducendo al contempo costi e tempi di risposta. Il software DRAGEN può contribuire all'ampliamento delle funzionalità di ricerca in diversi modi:

- 1. Stando al passo con i sistemi NovaSeq™ X Series, NextSeq™ 1000 System e NextSeq 2000 System.** Il software DRAGEN integrato può eseguire più applicazioni contemporaneamente (quattro applicazioni simultanee con un massimo di una conversione BCL e altre tre pipeline a scelta) per ogni cella a flusso in un'unica corsa.
- 2. Aumentando la capacità.** Nei periodi in cui i carichi di lavoro con elevati volumi di campioni sono più ingenti, i laboratori possono utilizzare l'analisi secondaria DRAGEN sulle applicazioni Illumina Connected

Analytics o DRAGEN su BaseSpace Sequence Hub sfruttando la capacità aggiuntiva sul cloud (figura 3).

- 3. Ampliando le operazioni.** Una sola istanza di DRAGEN consente di eseguire una vasta gamma di pipeline e di elaborare tutti i tipi di campioni supportati. Grazie alla completezza e all'efficienza del software DRAGEN, gli utenti possono adeguare le operazioni ai volumi senza compromettere i tempi di risposta o la qualità dei risultati.
- 4. Agevolando la transizione ai genomi.** Le pipeline preconfigurate di DRAGEN consentono una facile transizione dai pannelli mirati agli esomi e ai genomi.
- 5. Con iniziative correlate alla genomica della popolazione su larga scala.** L'analisi secondaria DRAGEN offre un flusso di lavoro semplificato per l'analisi di coorti su larga scala, con l'utilizzo congiunto di più pipeline per identificare variazioni genetiche in modo altamente accurato. DRAGEN gVCF Genotyper consente l'aggregazione da migliaia a milioni di file nel formato di identificazione di varianti del genoma (gVCF, genome Variant Call Format) e incorpora nuovi batch senza necessità di rielaborare i batch esistenti. La compressione ORA fa risparmiare sui costi di archiviazione.
- 6. Con applicazioni di sequenziamento profondo.** L'analisi secondaria DRAGEN supporta l'analisi altamente efficiente del sequenziamento a elevata profondità per una copertura media pari a oltre 300× per i genomi e 1.000× per gli esomi. Le capacità di sequenziamento profondo sono preziose per applicazioni come la ricerca oncologica e gli studi sulle malattie genetiche rare.

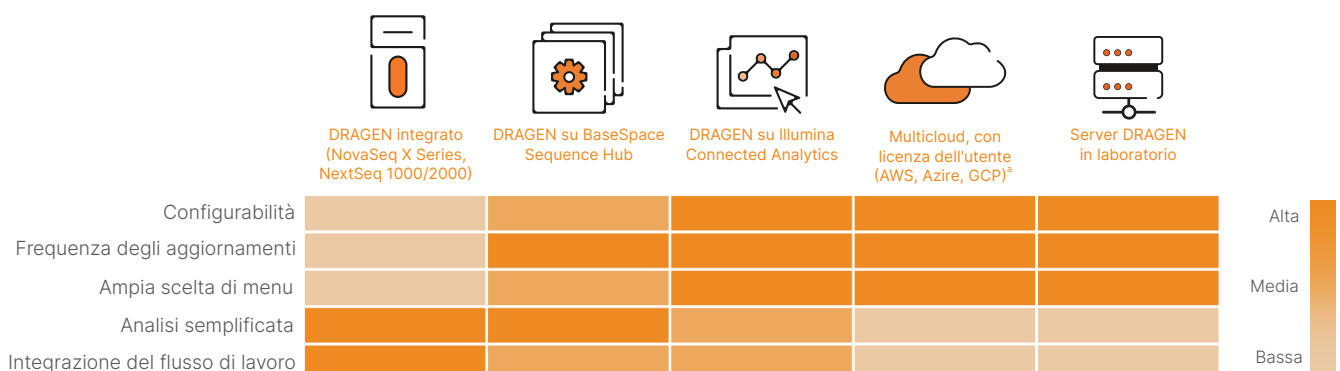


Figura 3: opzioni di accesso alle pipeline di DRAGEN con funzioni pensate per soddisfare le esigenze di ogni laboratorio per l'analisi dei dati di NGS.

a. Contattare il rappresentante Illumina per informazioni sull'accesso da Amazon Web Service (AWS), Azure o Google Cloud Platform (GCP, accesso anticipato).

Accessibilità da diverse piattaforme

Alla suite delle pipeline di DRAGEN si accede tramite soluzioni disponibili in laboratorio, integrate sullo strumento o sul cloud, così i laboratori possono scegliere quella che meglio soddisfa le loro esigenze (figura 3).

Server DRAGEN in laboratorio

Il server DRAGEN presente in laboratorio si appoggia a una soluzione di archiviazione locale per raccogliere e memorizzare i dati di NGS. Una volta trasferiti dallo strumento di sequenziamento all'archivio locale mediante una connessione di rete locale, i dati di sequenziamento non elaborati vengono inviati al server DRAGEN per l'esecuzione del flusso di lavoro selezionato. Dopo l'analisi, il software scrive i file di output generati nella posizione di archiviazione locale. Il server DRAGEN in laboratorio:

- supporta la configurazione flessibile delle relative funzioni tramite un'interfaccia a riga di comando;
- sostituisce fino a 30 istanze di calcolo tradizionali;
- elabora i dati di NGS per un intero genoma umano con una copertura pari a 40× in circa 35 minuti.

DRAGEN incluso in NovaSeq X Series

NovaSeq X Series integra l'analisi secondaria DRAGEN, una soluzione accurata, automatizzata e ottimizzata pensata per gestire gli straordinari volumi di dati generati da NovaSeq X Series. La suite integrata di software DRAGEN offre le funzionalità di identificazione di varianti e compressione ORA con le applicazioni di NGS (tabella 1), inclusi conversione BCL, linea germinale, somatica, arricchimento, RNA e metilazione. La suite DRAGEN integrata:

- esegue in parallelo diverse pipeline di analisi secondaria;
- esegue contemporaneamente fino a quattro applicazioni per ogni cella a flusso in una sola corsa;
- comprime i dati fino a 5× senza perdite, consentendo di risparmiare sui costi di archiviazione;
- permette di risparmiare sui costi di analisi, che nell'arco di cinque anni possono superare il prezzo d'acquisto di NovaSeq X System.

DRAGEN integrato in NextSeq 1000 System e NextSeq 2000 System

NextSeq 1000 System e NextSeq 2000 System contano sull'integrazione del software DRAGEN per un'esecuzione accurata e rapida dell'analisi secondaria. Si accede al software tramite un'interfaccia grafica intuitiva, che consente agli utenti esperti e non esperti di svolgere le analisi necessarie e ottenere rapidamente i risultati. Il software DRAGEN integrato offre pipeline selezionate per coprire una serie di applicazioni comuni di NGS (tabella 1) e include la premiata analisi ML e con riferimento (grafico) multigenomico per l'identificazione di alta qualità di varianti. La suite DRAGEN integrata:

- è più accurata di ogni altro sistema di sequenziamento da banco con analisi secondaria DRAGEN integrata;
- consente di accedere a pipeline informatiche di 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 un'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 degli 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, basato su 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;
- offre l'accesso a risorse di calcolo efficaci senza investimento di capitale per l'acquisto di infrastrutture aggiuntive.

ILLUMINA Connected Analytics

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. Su ILLUMINA Connected Analytics è possibile accedere alla suite di analisi secondaria DRAGEN, sotto forma di pipeline preconfezionate o singoli strumenti da incorporare in pipeline personalizzate.

Riepilogo

L'analisi secondaria DRAGEN è una suite efficace di strumenti software che consente di analizzare i dati di NGS in modo accurato, esaustivo ed efficiente. I laboratori possono scegliere tra le diverse opzioni di distribuzione del software DRAGEN quella più adatta al tipo e alle dimensioni dei loro progetti. Inoltre, gli utenti sono liberi di combinare varie opzioni di distribuzione per soddisfare al meglio le proprie esigenze a livello di prestazioni e flusso di lavoro. Di pari passo con i continui progressi della tecnologia di NGS, gli aggiornamenti tempestivi apportati all'analisi secondaria DRAGEN assicurano le migliori prestazioni possibili delle pipeline attuali, mentre nuove pipeline continuano a essere aggiunte man mano che le applicazioni diventano disponibili.

Maggiori informazioni

[Analisi secondaria DRAGEN](#)

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

[Contatti](#)



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-00680 ITA v10.0

Bibliografia

1. The Food and Drug Administration. Truth Challenge V2: Calling Variants from Short and Long Reads in Difficult-to-Map Regions. <https://precision.fda.gov/challenges/10/results>. Consultato il 14 marzo 2022.
2. Catreux S, Jain V, Murray L, et al. DRAGEN sets new standard for data accuracy in PrecisionFDA benchmark data. Optimizing variant calling performance with Illumina machine learning and DRAGEN graph. Sito Web Illumina. <https://www.illumina.com/science/genomics-research/articles/dragen-shines-again-precisionfda-truth-challenge-v2.html>. Pubblicato il 12 gennaio 2022. Consultato l'11 luglio 2024.
3. Mehio R, Ruehle M, Catreux S, et al. DRAGEN wins at PrecisionFDA Truth Challenge V2 showcase accuracy gains from alt-aware mapping and graph reference genomes. Sito Web Illumina. illumina.com/science/genomics-research/articles/dragen-wins-precisionfda-challenge-accuracy-gains.html. Pubblicato il 9 novembre 2020. Consultato l'11 luglio 2024.
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. Pubblicato il 23 ottobre 2017. Consultato l'11 luglio 2024.
6. San Diego Union Tribune. Rady Children's Institute sets Guinness world record. <https://www.sandiegouniontribune.com/2018/02/12/rady-childrens-institute-sets-guinness-world-record/>. Pubblicato il 12 febbraio 2018. Consultato l'11 luglio 2024.
7. Betschart RO, Thiéry A, Aguilera-Garcia D, et al. [Comparison of calling pipelines for whole genome sequencing: an empirical study demonstrating the importance of mapping and alignment](#). *Sci Rep.* 2022;12(1):21502. Pubblicato il 13 dicembre 2022. doi:10.1038/s41598-022-26181-3
8. Gross A, Maciuga S, Cox A, et al. Accurate and Efficient Calling of Small and Large Variants from PopGen data sets Using the DRAGEN Bio-IT Platform. Sito Web Illumina. www.illumina.com/science/genomics-research/articles/popgen-variant-calling-with-dragen.html. Pubblicato il 24 maggio 2021. Consultato il 14 marzo 2022.