

Sequenziamento completo dell'intero genoma con Illumina Complete Long Read Prep, Human

Prestazioni eccezionali in
regioni difficili da mappare

illumina[®]

Introduzione

Il sequenziamento di nuova generazione (NGS, Next-Generation Sequencing) consente agli scienziati di decifrare il genoma per ottenere una comprensione più approfondita della biologia. La comprovata chimica di sequenziamento mediante sintesi (SBS, Sequencing By Synthesis) di Illumina combinata con la premiata analisi secondaria DRAGEN™ fornisce dati di sequenziamento dell'intero genoma (WGS, Whole-genome Sequencing) con un'accuratezza eccezionale.¹ In alcuni casi, una piccola porzione del genoma può trarre vantaggio da lunghezze di lettura maggiori che consentono una risoluzione e una mappatura più accurate di queste regioni difficili. Tuttavia, molte soluzioni di sequenziamento per letture lunghe hanno flussi di lavoro complessi, elevati requisiti di input di DNA e risultati altamente variabili.²⁻⁵

L'Illumina Complete Long Reads semplifica il sequenziamento di letture lunghe e lo rende accessibile agli scienziati genomici. L'Illumina Complete Long Read Prep, Human è il primo prodotto basato su questa nuova chimica per letture lunghe. Il saggio WGS a elevate prestazioni utilizza un flusso di lavoro NGS standard per generare sequenze di letture lunghe contigue sul NovaSeq™6000 System e sulla NovaSeq X Series (Figura 1). L'efficiente protocollo di preparazione delle librerie può essere completato in un giorno, è facile da ampliare per gli studi a elevata processività e richiede appena 10 ng di input di DNA senza che sia necessario eseguire estrazioni specializzate, frammentazione (shearing) o selezione della dimensione.

Questa nota tecnica dimostra le eccezionali prestazioni di Illumina Complete Long Read Prep, Human con un'ampia gamma di input di DNA, qualità del DNA e tipi di campione per WGS altamente accurati e completi per l'analisi delle varianti della linea germinale.

Metodi

Il campione di riferimento HG002/NA24385 è stato ottenuto dal Genome in a Bottle consortium (GIAB) come DNA genomico (gDNA, genomic DNA) purificato. Separatamente, sono stati ottenuti campioni di sangue e saliva e il gDNA è stato estratto utilizzando kit disponibili in commercio con estrazione regolare o ad alto peso molecolare, seguendo le istruzioni del produttore.

Preparazione delle librerie

Le librerie "marcate" sono state preparate da quantità variabili di gDNA di input utilizzando Illumina Complete Long Read Prep, Human (Illumina, n. di catalogo 20089108). In breve, i frammenti lunghi di DNA a singola molecola vengono marcati enzimaticamente con pattern unici (o landmark). Le librerie non marcate, utilizzate durante l'analisi per produrre letture contigue lunghe che rappresentano il frammento monomolecolare originale, sono state preparate utilizzando Illumina DNA PCR-Free Prep, Tagmentation (Illumina, n. di catalogo 20041794).

Sequenziamento

Le librerie marcate e non marcate sono state sequenziate durante corse separate sul NovaSeq 6000 System a una lunghezza di lettura di 2×150 bp.

Analisi dei dati

Il raggruppamento e il rendering delle letture marcate sono stati eseguiti utilizzando l'Illumina Complete Long Read WGS App su BaseSpace™ Sequence Hub. Per l'analisi e la visualizzazione dei dati è stato utilizzato Interactive Genomics Viewer (IGV).



Figura 1: flusso di lavoro di Illumina Complete Long Read Prep, Human. Illumina Complete Long Read Prep, Human offre un flusso di lavoro ottimizzato costituito da quattro fasi, ovvero, estrazione del DNA, preparazione delle librerie, sequenziamento e analisi dei dati.

a. Sebbene sia possibile utilizzare un input minimo di 10 ng, Illumina raccomanda di utilizzare 50 ng di DNA.

Dati di alta qualità in regioni difficili

ILLUMINA Complete Long Read Prep, Human fornisce una copertura nelle regioni in cui le letture brevi non possono essere mappate in modo inequivocabile. Ciò consente il sequenziamento contiguo di geni e pseudogeni associati alla malattia con problemi di mappatura noti solo con letture brevi, tra cui *STRC*, (Figura 2A), *SEMG1* (Figura 2B) e *SULTIA1* (Figura 2C).

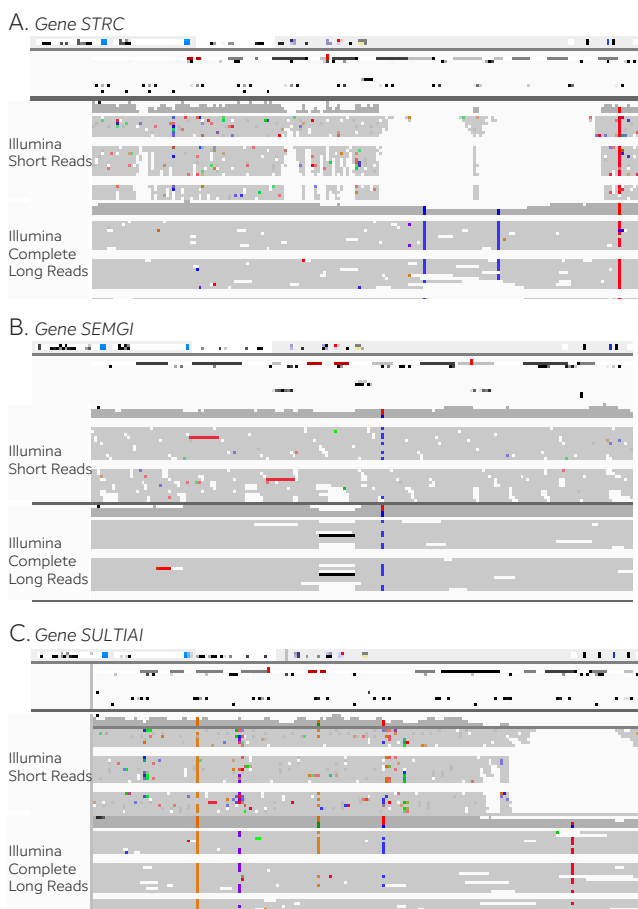


Figura 2: la copertura completa consente il rilevamento accurato di più tipi di varianti in regioni difficili. Illumina Complete Long Read Prep, Human consente il sequenziamento accurato dei geni associati alla malattia in regioni genomiche difficili, tra cui (A) *STRC*, (B) *SEMG1* e (C) *SULTIA1*.

Inoltre, Illumina Complete Long Read Prep, Human fornisce blocchi di fase significativamente più grandi rispetto alla dimensione genica tipica. Ciò consente una copertura uniforme in tutto il locus dell'antigene leucocitario umano (HLA, Human Leukocyte Antigen), una regione densamente polimorfica con elevati livelli di omologia delle sequenze, pseudogeni, variabilità e numerosi alleli. Sebbene sia difficile eseguire un sequenziamento accurato, Illumina Complete Long Read Prep, Human fornisce una copertura uniforme in tutta la regione HLA e consente la determinazione delle fasi del genoma per visualizzare e distinguere chiaramente gli alleli tra i cromosomi materni e paterni (Figura 3).



Figura 3: la copertura uniforme consente una chiara determinazione delle fasi degli alotipi. Illumina Complete Long Read Prep, Human fornisce una copertura completa nella regione HLA e consente una determinazione delle fasi chiara e accurata degli alleli all'interno della regione HLA.

Per una valutazione più completa dell'accuratezza dell'identificazione di varianti, Illumina Long Read Prep, Human è stato messo a confronto con una piattaforma di lettura lunga di terze parti e di lettura breve Illumina in PrecisionFDA Truth Challenge V2. Sponsorizzata da PrecisionFDA, dal GIAB e dal National Institute of Standards and Technology (NIST), questa sfida è stata lanciata per valutare le prestazioni della pipeline di identificazione di varianti piccole su un quadro di riferimento comune, con particolare attenzione alle regioni "difficili da mappare". Utilizzando questo set di dati, Illumina Complete Long Read Prep, Human dimostra un'accuratezza eccezionale e supera la piattaforma di lettura lunga alternativa con un punteggio F1 pari al 99,87% (Figura 4).

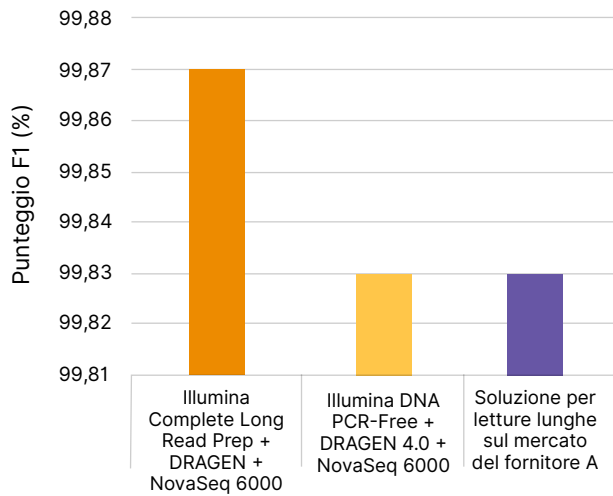


Figura 4: un nuovo standard di accuratezza. Illumina Complete Long Read Prep, Human fornisce un'accuratezza senza precedenti per l'identificazione di varianti, secondo quanto misurato dal punteggio F1 (%), generando precisione e richiamo per il WGS.¹

Solide prestazioni su tutte le quantità di input di DNA

Le prestazioni di Illumina Complete Long Read Prep, Human sono state valutate per una serie di quantità di input di DNA. Le librerie sono state preparate da 10, 25, 100, 600 e 1.200 ng di DNA in triplicato. I risultati del sequenziamento hanno dimostrato che le librerie preparate da appena 5 ng di input mostravano metriche di alta qualità, tra cui N50 e N50 dei blocchi di determinazione delle fasi (Figura 5). Anche se Illumina raccomanda un input di 50 ng di DNA, è possibile utilizzare input inferiori.

Supporto per diverse qualità di DNA

Per simulare il DNA danneggiato, il gDNA di alta qualità è stato sottoposto a frammentazione (shearing) utilizzando un sonificatore a ultrasuoni Covaris per uno o tre secondi. La qualità del DNA è stata valutata utilizzando un analizzatore di frammenti (Figura 6). Le librerie sono state preparate utilizzando input di gDNA con frammentazione e Illumina Complete Long Read Prep, Human. I risultati del sequenziamento hanno mostrato che il DNA con frammentazione per un secondo generava dati di alta qualità e consentiva l'identificazione di varianti accurata (Figura 7). Inoltre, l'input di gDNA sottoposto a cicli crescenti di congelamento e scongelamento ha prodotto dati di alta qualità (Figura 8).

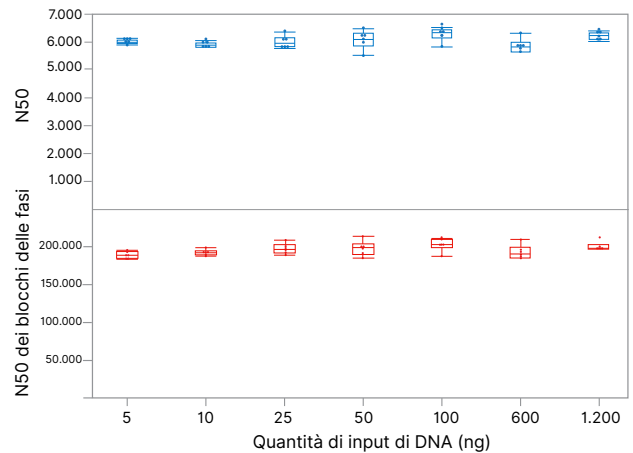


Figura 5: prestazioni di alta qualità su un'ampia gamma di input di DNA. Illumina Complete Long Read Prep, Human produce librerie di alta qualità preparate da input di DNA da 5 ng a 1.200 ng (in triplicato) che generano una qualità dei dati simile per N50 e N50 di blocchi di fase. N50 è definito come la lunghezza della sequenza del contig più breve al 50% della lunghezza dell'assemblaggio totale; può essere utilizzato anche come misura dei blocchi di fase.

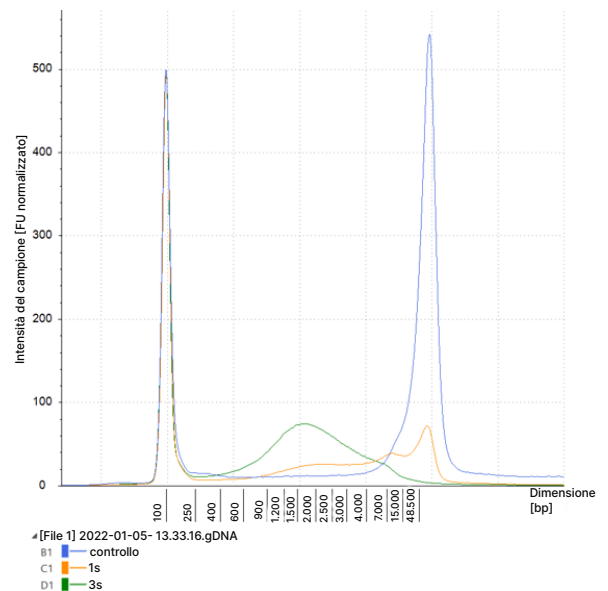


Figura 6: qualità del DNA con frammentazione. L'analisi del DNA su un analizzatore di frammenti mostra una qualità ridotta con un maggiore tempo di frammentazione. I campioni di DNA sono stati sottoposti a frammentazione Covaris per 1 o 3 secondi e la qualità è stata misurata utilizzando un analizzatore di frammenti.



Figura 7: metriche di sequenziamento del DNA con frammentazione. Il DNA con frammentazione per 1 e 3 secondi è stato utilizzato come input per Illumina Complete Long Read Prep, Human. Le librerie risultanti hanno generato dati di sequenziamento a lettura lunga di alta qualità, secondo quanto misurato mediante N50 e N50 di blocchi di fase.

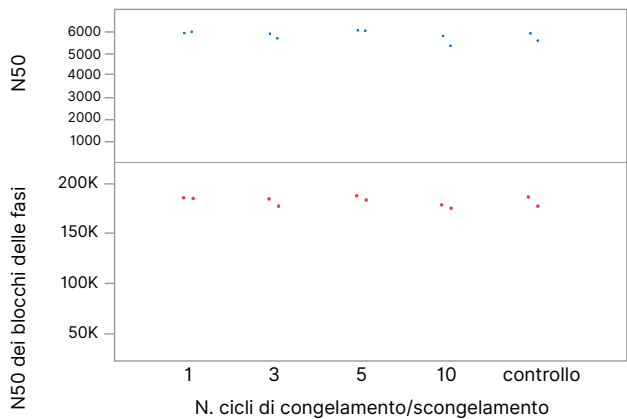


Figura 8: metriche di sequenziamento del DNA congelato-scongelato. Il DNA sottoposto a 1, 3, 5 e 10 cicli di congelamento-scongelo è stato utilizzato come input per Illumina Complete Long Read Prep, Human. Le librerie risultanti hanno generato dati di sequenziamento a lettura lunga di alta qualità, secondo quanto misurato mediante N50 e N50 di blocchi di fase.

Supporto flessibile per diversi tipi di campioni

Illumina Complete Long Read Prep, Human supporta diversi tipi di campioni, tra cui sangue e saliva. Dopo l'estrazione tramite un kit standard o ad alto peso molecolare (HMW, High Molecular Weight), la qualità del DNA è stata valutata utilizzando un analizzatore di frammenti (Figura 9). I risultati del sequenziamento hanno mostrato che il DNA derivato dal sangue e dalla saliva generava dati di alta qualità (Figura 10). Non vi era alcuna differenza di prestazioni tra il DNA estratto con un kit standard e quello estratto con un kit HMW (Figura 10).

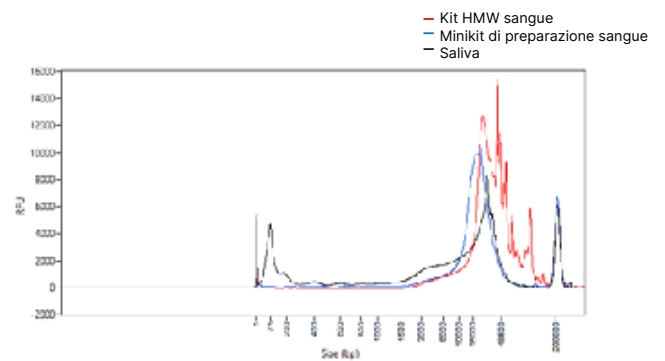


Figura 9: qualità del DNA di sangue e saliva. Il DNA estratto dal sangue (HMW, mini) o dalla saliva è stato analizzato su un analizzatore di frammenti.

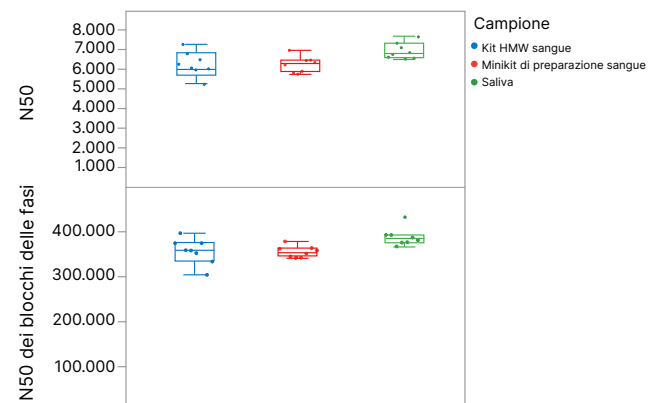


Figura 10: prestazioni di alta qualità con diversi tipi di campioni. Le librerie Illumina Complete Long Read Prep, Human generate dal sangue (HMW, mini) e dalla saliva forniscono dati di alta qualità, secondo quanto misurato da N50 e N50 di blocchi di fase.

Riepilogo

La chimica del sequenziamento a lettura lunga fornisce informazioni aggiuntive per le regioni del genoma che sono difficili da risolvere con l'NGS a lettura breve. Grazie a Illumina Complete Long Reads, il sequenziamento di letture lunghe risulta accessibile e ottimizzato per gli scienziati genomici. Illumina Complete Long Read Prep, Human, il primo prodotto realizzato su questa nuova chimica, offre un flusso di lavoro e una sinergia semplificati e familiari con la comprovata chimica SBS Illumina e l'analisi DRAGEN. Questa chimica altamente innovativa consente di eseguire il sequenziamento a lettura breve e lunga su un singolo strumento, rendendo accessibile l'NGS a lettura lunga ai laboratori genomici. Illumina Complete Long Read Prep, Human dimostra solide prestazioni con DNA di quantità e qualità variabili da diverse fonti di campioni. Il risultato è una soluzione WGS umana altamente scalabile e accurata.

Maggiori informazioni

[Illumina Complete Long Read Prep, Human](#)

Bibliografia

1. Mehio R, Ruehle M, Catreux S, et al. DRAGEN Wins at Precision- FDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. [illumina.com/science/genomics-research/dragen-wins-precisionfda-challenge-showcase-accuracy-gains.html](https://www.illumina.com/science/genomics-research/dragen-wins-precisionfda-challenge-showcase-accuracy-gains.html). Consultato il 12 gennaio 2023.
2. Pacific Biosciences. Preparing DNA for PacBio HiFi sequencing—Extraction and quality control. [pacb.com/wp-content/uploads/Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf](https://www.pacb.com/wp-content/uploads/Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf). Pubblicato nel 2022. Consultato il 12 gennaio 2023.
3. Pacific Biosciences. Preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0. [pacb.com/wp-content/uploads/Procedure-checklist-Preparing-whole-genome-and-metagenome-libraries-using-SMRTbell-prep-kit-3.0.pdf](https://www.pacb.com/wp-content/uploads/Procedure-checklist-Preparing-whole-genome-and-metagenome-libraries-using-SMRTbell-prep-kit-3.0.pdf). Pubblicato nel 2022. Consultato il 12 gennaio 2023.
4. Oxford Nanopore Technologies. Ligation Sequencing Kit. [store.nanoporetech.com/us/ligation-sequencing-kit110.html](https://www.store.nanoporetech.com/us/ligation-sequencing-kit110.html). Consultato il 12 gennaio 2023.
5. Pacific Biosciences. Low Yield Troubleshooting Guide. [pacb.com/wp-content/uploads/Guide-Low-Yield-Troubleshooting.pdf](https://www.pacb.com/wp-content/uploads/Guide-Low-Yield-Troubleshooting.pdf). Pubblicato nel 2018. Consultato il 12 gennaio 2023.



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-01421 ITA v1.0