

illumina TruPath™ Genome

Il più semplice flusso di lavoro per il sequenziamento dell'intero genoma



Flusso di lavoro dal campione ai dati estremamente semplice, con preparazione manuale di 10 minuti

TAGCA
CGTAG
ATAGC

Risoluzione delle regioni del genoma difficili da mappare con letture brevi e dati di prossimità utilizzando NovaSeq™ X Series



Acquisizione di informazioni su lunghe distanze grazie al rilevamento migliorato delle varianti strutturali e alla determinazione delle fasi su tratti ultralunghi

Accessibilità rivoluzionaria del genoma

Il sequenziamento di nuova generazione (NGS, Next-Generation Sequencing) a lettura breve è il metodo principale utilizzato dai ricercatori per saggiare la maggior parte del genoma. Tuttavia, alcune regioni e alcuni tipi di varianti rimangono difficili da caratterizzare, come quelli contraddistinti da estrema ripetitività, elevata omologia delle sequenze o complessità strutturale. Recenti studi hanno dimostrato che queste regioni ospitano varianti che possono contribuire alle malattie genetiche nell'uomo,^{1,2} il che rende la loro decifrazione della massima importanza ai fini della ricerca delle cause alla base delle malattie genetiche.

Metodi alternativi, come il sequenziamento a lettura lunga, offrono una risoluzione migliorata, la quale consente a sua volta di ottenere maggiori informazioni su molte di queste varianti e regioni difficili da mappare. I metodi a lettura lunga, tuttavia, tendono ad avere flussi di lavoro complessi e problemi di accuratezza radicati che portano a risultati variabili.

Illumina TruPath Genome con [tecnologia di lettura basata sulla mappatura](#) di prossimità cambia il paradigma del flusso di lavoro per il sequenziamento dell'intero genoma (WGS, Whole Genome Sequencing). Questo saggio innovativo cattura lunghi frammenti di DNA con una semplice fase manuale di preparazione del saggio della durata di 10 minuti.

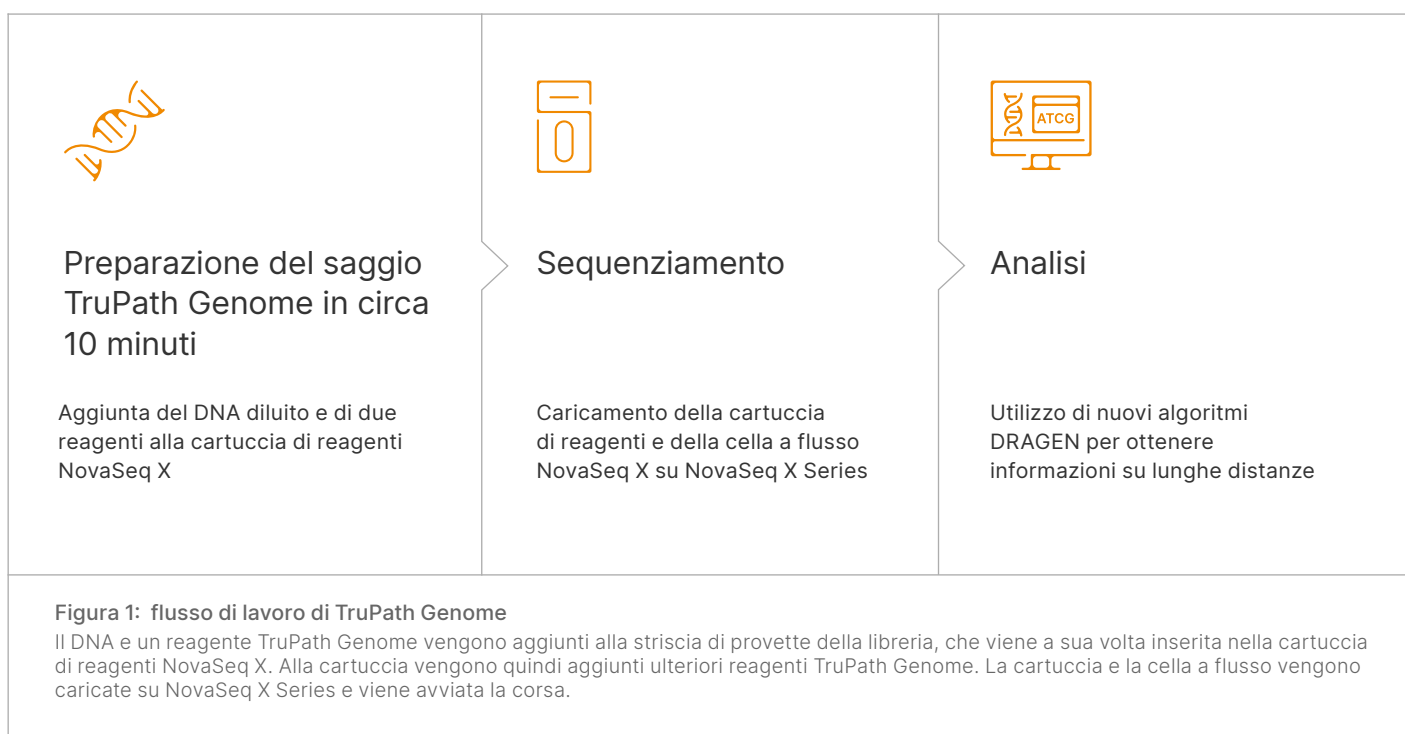
Il sequenziamento avviene sul sistema NovaSeq X Series esistente, con l'innovativa analisi secondaria DRAGEN™. Il risultato è un genoma accurato e completo, che consente ai ricercatori di generare informazioni che coprono distanze maggiori rispetto alle lunghezze di lettura tipicamente ottenute con i metodi a lettura lunga.

Con TruPath Genome, è ora possibile sfruttare la semplicità e la precisione del sequenziamento a lettura breve per risolvere le regioni del genoma difficili da mappare, migliorare il rilevamento delle varianti strutturali (SV, Structural Variant) e generare identificazioni di varianti e letture con determinazione delle fasi.

Semplicità senza precedenti del flusso di lavoro

TruPath Genome semplifica notevolmente l'implementazione del WGS e riduce gli interventi manuali a 10 minuti ([Figura 1](#)). L'affidabile chimica della trasposasi Illumina elimina le fasi tradizionali di preparazione delle librerie. Gli algoritmi bioinformatici avanzati sfruttano le informazioni di prossimità dei nanopozzetti insieme a dati standard paired-end e a lettura breve* per risolvere le relazioni tra le varianti separate anche da milioni di basi.

* Per dati standard paired-end e a lettura breve si intendono i dati generati utilizzando il sequenziamento dell'intero genoma basato su preparazione manuale delle librerie e metodi standard di sequenziamento mediante sintesi (SBS, Sequencing By Synthesis).



Campioni

Il DNA a doppio filamento (dsDNA, double-stranded DNA) purificato viene estratto con kit commerciali standard o ad alto peso molecolare (HMW, High Molecular Weight). I 350 ng di DNA di input raccomandati vengono diluiti con il tampone di tagmentazione TruPath e aggiunti alla striscia di provette della libreria. Il DNA di input può essere isolato da linee cellulari o sangue intero. Il DNA isolato da campioni fissati in formalina e inclusi in paraffina (FFPE, Formalin-Fixed, Paraffin-Embedded) e il DNA libero circolante (cfDNA, cell-free DNA) non sono supportati. Per prestazioni ottimali, si raccomanda di estrarre il DNA utilizzando un metodo in grado di preservare i frammenti grandi, di modo che il 40% di essi abbia una lunghezza superiore a 60 kb.³

Per indicazioni sull'utilizzo di ulteriori tipi di campione, sui diversi gradi di qualità degli stessi e sulle quantità di input ottimali, leggere la [nota tecnica Prestazioni di TruPath Genome con campioni di tipi e qualità variabili](#).

Sequenziamento

La striscia di provette della libreria con il DNA diluito e due reagenti TruPath Genome vengono aggiunti alla cartuccia di reagenti NovaSeq X e caricati su NovaSeq X Series con una cella a flusso NovaSeq X C2 o C8 per il sequenziamento di due o otto genomi umani rispettivamente.

I nanopozzetti sulla superficie della cella a flusso sono sottoposti a priming con trasposoni Illumina.[†] Il DNA viene introdotto nella cella a flusso e qui catturato dai trasposoni in numerosi pozzetti, creando uno schema di nanopozzetti seminati simile a una costellazione (Figura 2A). I trasposoni scindono il DNA catturato e attaccano i frammenti alla superficie della cella a flusso in un processo detto tagmentazione[‡] (Figura 2B). Una fase di lavaggio rimuove il DNA non legato e i trasposoni.

All'interno dei nanopozzetti, il DNA tagmentato viene utilizzato per creare dei cluster[§] per il sequenziamento. Il DNA dello stesso frammento lungo di origine andrà a formare dei cluster nei nanopozzetti vicini. Il sequenziamento avviene su NovaSeq X Series con la comprovata chimica di sequenziamento mediante sintesi XLEAP (XLEAP-SBS™) di Illumina e corsa di 2 × 150 bp.

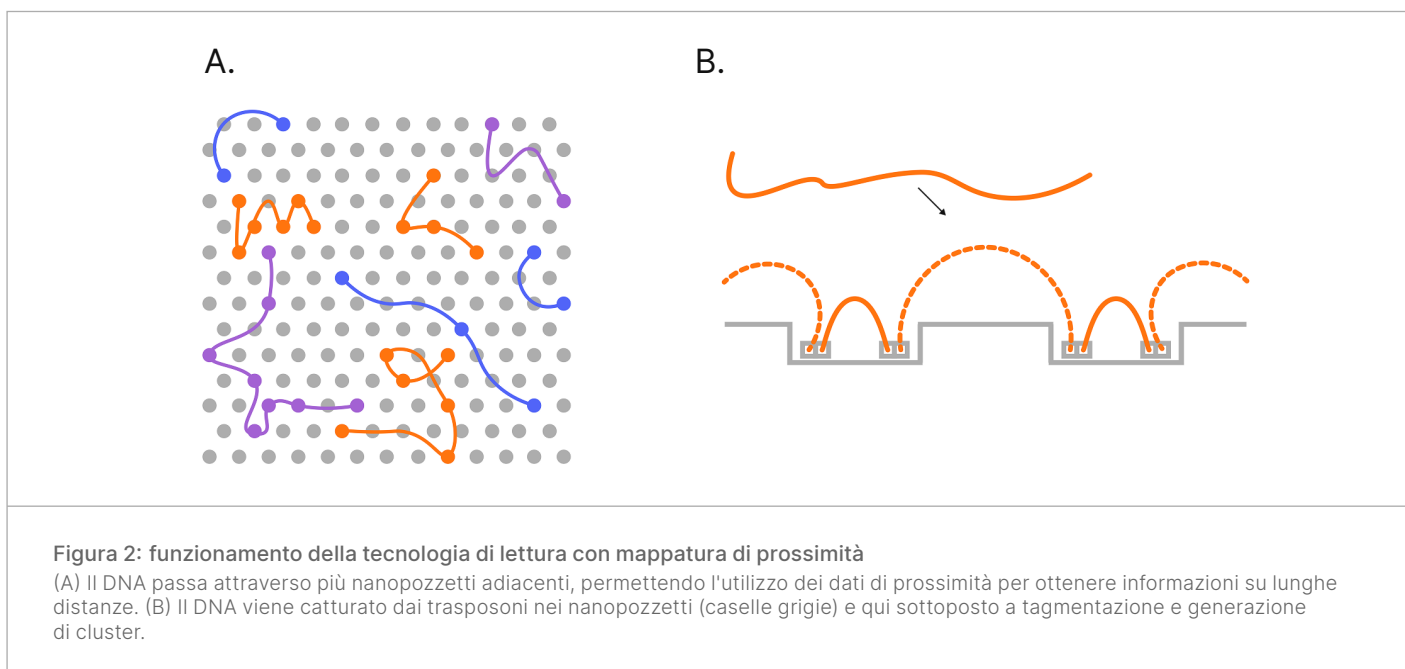
Analisi

Al termine del sequenziamento, la pipeline DRAGEN™ Germline utilizza un nuovo algoritmo che combina i dati di sequenziamento con le informazioni di prossimità dei nanopozzetti per allineare il DNA a un genoma di riferimento. Disporre delle informazioni di prossimità consente di assegnare correttamente le letture che in precedenza non erano mappate sul genoma di riferimento o che potevano essere allineate solo in modo scarsamente affidabile.

[†] I trasposoni sono complessi di trasposasi del DNA che si presentano sotto forma di dimeri.

[‡] La tagmentazione è il processo di taglio di un frammento di DNA e aggiunta di una sequenza adattatore (tagging) utilizzando un trasposone.

[§] Un cluster corrisponde a una regione di DNA amplificata su una cella a flusso che sarà sequenziata.



Con questa modalità, le letture brevi standard possono essere associate tra loro in maniera probabilistica, così da consentire il rilevamento completo delle varianti, tra cui varianti piccole [varianti a singolo nucleotide (SNV, Single-Nucleotide Variant) e inserzioni/delezioni (indel)], varianti grandi (oltre 50 bp) e regioni omologhe note, e la determinazione delle fasi per le identificazioni di varianti piccole separate anche da milioni di coppie di basi.

Genoma completo con letture brevi altamente accurate

Il metodo standard di sequenziamento mediante sintesi (SBS, Sequencing By Synthesis) si è dimostrato in grado di rilevare con un alto livello di precisione le varianti piccole [ad esempio, SNV, indel e variazioni del numero di copie (CNV, Copy Number Variations)], ma incontra delle difficoltà con regioni e tipi di varianti specifici, come le varianti strutturali. TruPath Genome aggiunge nuove funzionalità al sequenziamento a lettura breve al fine di ottenere un genoma più completo. Ora le letture brevi possono essere utilizzate per analizzare regioni del genoma precedentemente considerate complesse e offrire maggiore accuratezza con ulteriori tipi di varianti.

Migliore copertura nelle regioni difficili da mappare

Le aree genomiche complesse, altamente polimorfiche o duplicate possono essere difficili da assemblare o allineare con un genoma di riferimento a causa della bassa mappabilità, sia per la ridotta copertura sia per la scarsa qualità della mappatura (MapQ). In precedenza, la copertura di queste regioni "oscure"⁴ causava bassi punteggi MapQ, lacune nei dati di sequenziamento del genoma e risoluzione complessiva limitata per l'identificazione di varianti. Per colmare queste lacune, TruPath Genome utilizza le informazioni di prossimità dei cluster adiacenti al fine di assegnare la posizione genomica corretta. In questo modo, le letture la cui mappatura mostrava delle ambiguità possono ora essere mappate in maniera altamente affidabile (Figura 3 e Figura 4).

Identificazione più accurata delle varianti

TruPath Genome è stato confrontato con il genoma HG002 e il truth set del consorzio Telomere-to-Telomere (T2T) T2T-Q100 v1.1 v0.019 per dimostrarne le prestazioni a livello di identificazione di varianti piccole, rilevamento di SV e sequenziamento con determinazione delle fasi. La maggiore risoluzione della mappatura consentita dalla tecnologia di lettura con mappatura di prossimità ha migliorato le prestazioni dell'identificazione di varianti piccole, in particolare nelle regioni del genoma tradizionalmente difficili da mappare, come quelle omologhe o ripetitive.

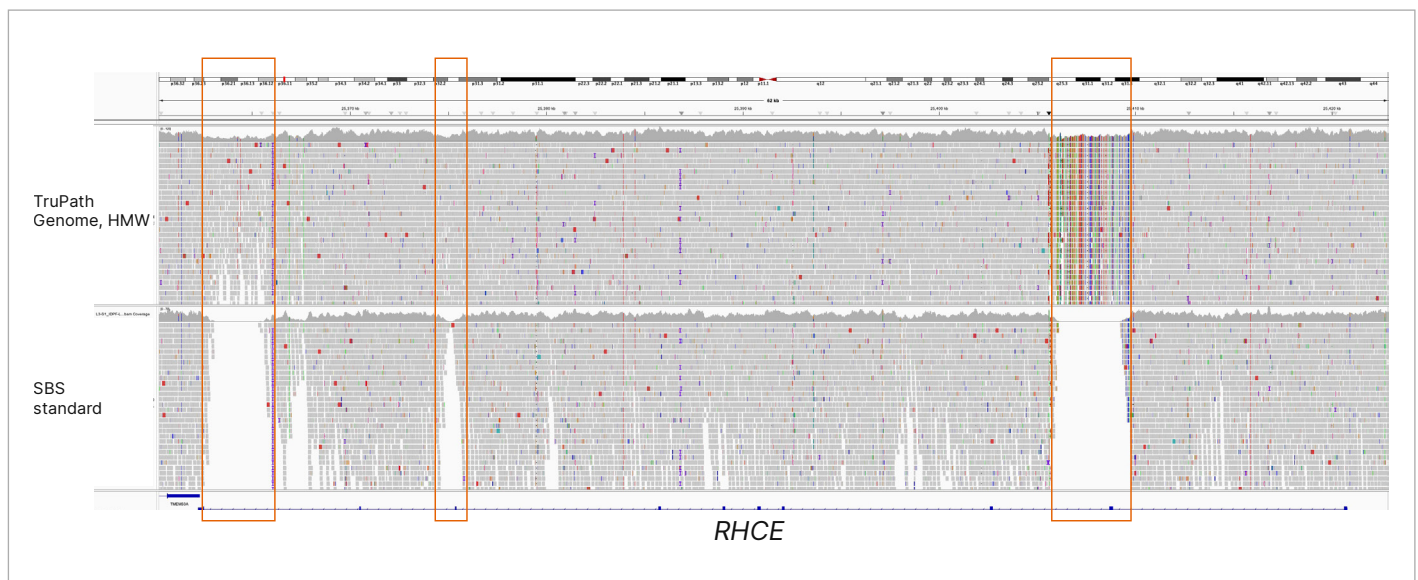


Figura 3: risoluzione efficace di geni difficili da mappare e clinicamente rilevanti con TruPath Genome

Recupero della copertura nel gene *RHCE*, che si trova nel locus del gruppo sanguigno Rh adiacente a *RHD* e condivide un'omologia della sequenza molto elevata con frequenti conversioni geniche e riarrangiamenti ibridi, rendendo ambigua in alcuni punti la mappatura a lettura breve. L'accurata genotipizzazione del gene *RHCE* è clinicamente importante per determinare gli antigeni Rh (C/c, E/e) in quanto le mancate corrispondenze possono aumentare il rischio di reazioni emolitiche trasfusionali e di malattia emolitica del feto/neonato (HDFN, Hemolytic Disease of the Fetus/Newborn). L'SBS standard è stato eseguito con Illumina DNA PCR-Free Prep.

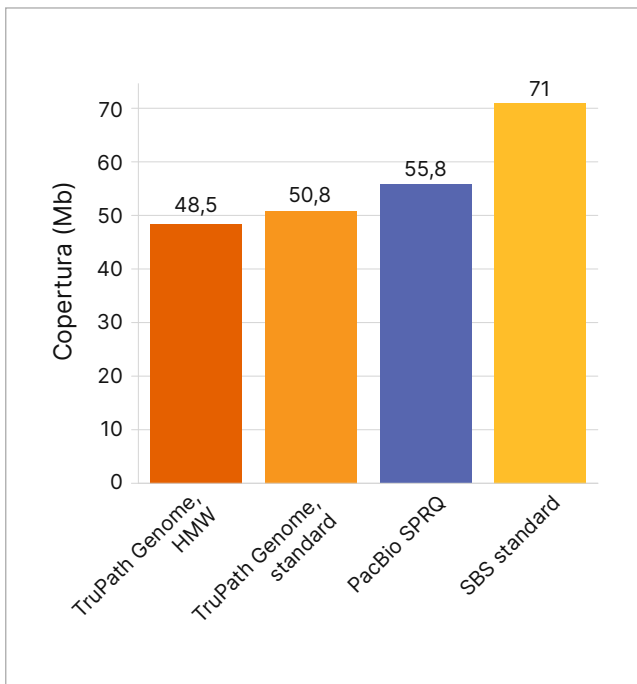


Figura 4: TruPath Genome migliora la copertura delle "regioni oscure" del genoma

TruPath Genome migliora la copertura delle "regioni oscure", definite regioni "dark-by-MAPQ" in Ebbert et al,¹ ossia quelle in cui il 90% delle letture ha una qualità di mappatura (MAPQ, Mapping Quality) inferiore a 10. TruPath Genome ha ridotto le dimensioni delle aree oscure del genoma grazie alla mappatura migliorata delle letture nelle regioni genomiche difficili. L'SBS standard è stato eseguito con Illumina DNA PCR-Free Prep.

Rispetto al sequenziamento e all'identificazione di varianti con tecnologie standard a lettura breve, TruPath Genome mostra maggiore accuratezza grazie alla riduzione sostanziale di falsi negativi (FN) e falsi positivi (FP) (Figura 5).

Con TruPath Genome, le prestazioni di identificazione di SV (oltre 50 bp) sono nettamente superiori rispetto a quelle osservate con i metodi standard a lettura breve. Il richiamo delle SV è passato dall'86% con l'SBS standard al 94% con la tecnologia di lettura con mappatura di prossimità (Figura 6).

Determinazione delle fasi su tratti ultralunghi

Il sequenziamento con determinazione delle fasi fornisce informazioni sugli aplotipi che consentono di distinguere gli alleli presenti sui cromosomi materno e paterno.⁸ Queste informazioni possono aiutare i ricercatori a risolvere varianti eterozigoti composte, aspetto fondamentale per l'interpretazione delle condizioni autosomiche recessive.

La tecnologia di lettura con mappatura di prossimità è ideale per la determinazione delle fasi del genoma umano. Il DNA catturato sulla cella a flusso produce blocchi con fasi determinate di lunghezza compresa tra centinaia di kilobasi e diverse megabasi (Figura 7 e Figura 8), contribuendo a generare maggiori informazioni sugli aplotipi e sugli eterozigoti composti.

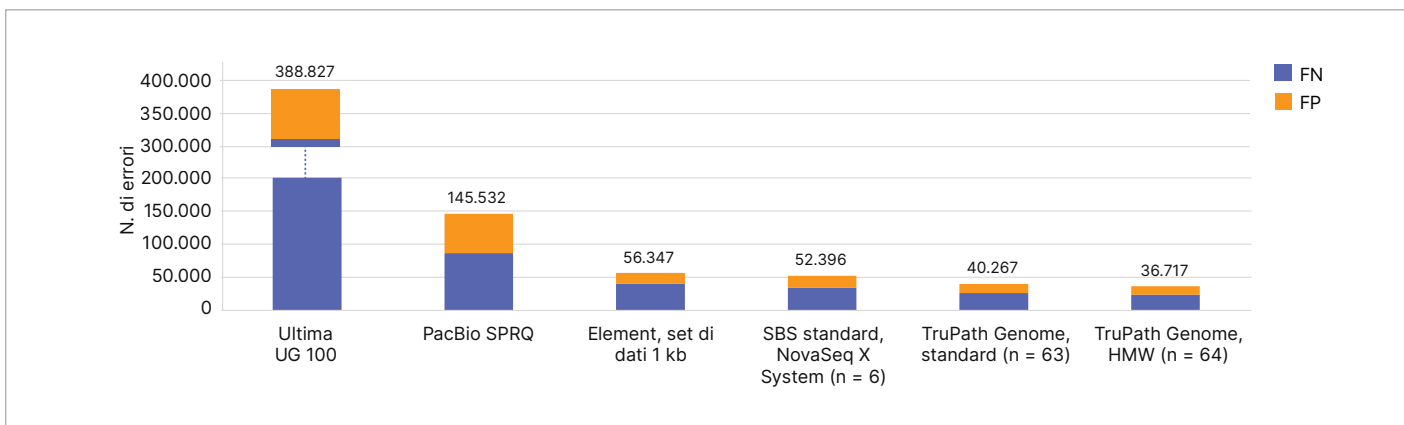


Figura 5: identificazione di varianti piccole più accurata con TruPath Genome

Le prestazioni dell'identificazione di varianti piccole con vari sistemi e saggi di NGS sono state confrontate con il truth set Genome in a Bottle NIST T2T-Q100 HG002 SV v1.1. I dati relativi a Ultima UG 100, PacBio con chimica SPRQ ed Element AVITI con set di dati di 1 kb sono stati ricavati da materiali pubblicati.⁵⁻⁷ I dati di SBS standard sono stati generati con le librerie Illumina DNA PCR-Free Prep sequenziate su NovaSeq X System con software v1.4 e cella a flusso 10B e analizzate con DRAGEN Germline v4.5.2 (sei replicati tecnici). I dati standard di TruPath Genome sono stati generati con DNA estratto mediante metodi standard, sequenziato su NovaSeq X System con software v1.4 e cella a flusso C8 e analizzato con DRAGEN Germline v4.5.2 (63 replicati tecnici). I dati HMW di TruPath Genome sono stati generati con DNA estratto mediante metodi HMW, sequenziato su NovaSeq X System con software v1.4 e cella a flusso C8 e analizzato con DRAGEN Germline v4.5.2 (64 replicati tecnici). FN, identificazioni di varianti false negative; FP, identificazioni di varianti false positive.

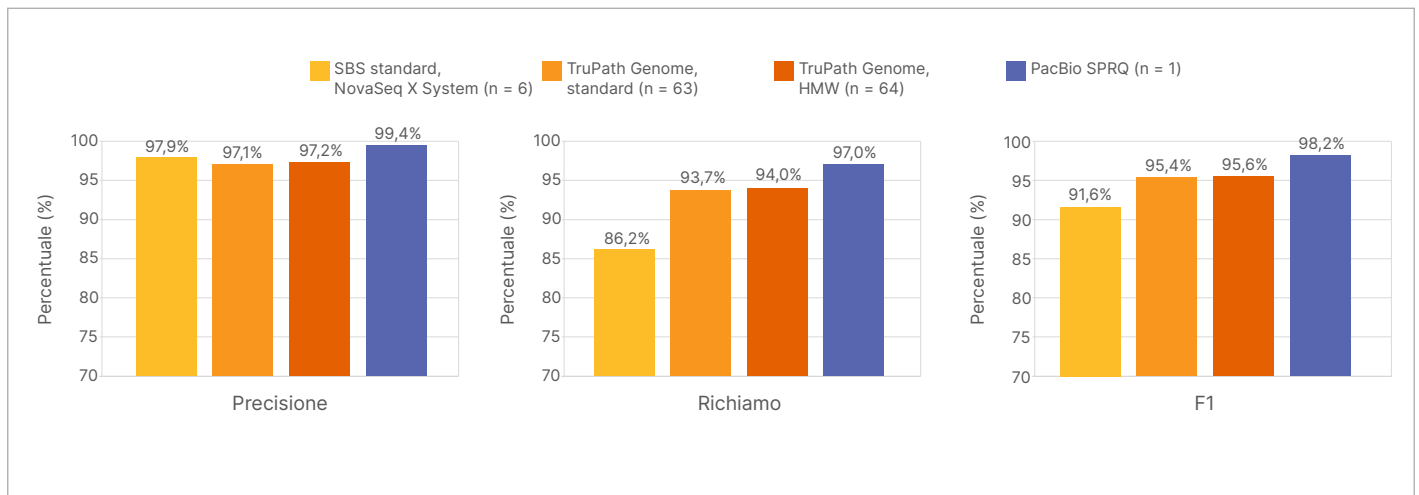


Figura 6: migliore identificazione di varianti strutturali con TruPath Genome

Confronto delle prestazioni relative alle varianti strutturali (SV, structural variant) di TruPath Genome e dell'SBS standard con DRAGEN Germline v4.5.2. L'analisi utilizza il truth set Genome in a Bottle NIST T2T-Q100 HG002 SV v1.1 con il file BED delle SV considerate affidabili. Il confronto è stato eseguito in conformità alle [linee guida Genome in a Bottle per il confronto delle varianti strutturali](#) utilizzando i comandi "bench" e "refine" di Truvari v4.2.2. I dati di SBS standard sono stati generati con le librerie Illumina DNA PCR-Free Prep sequenziate su NovaSeq X System con software v1.4 e cella a flusso 10B (sei replicati tecnici). I dati standard di TruPath Genome sono stati generati con DNA estratto mediante metodi standard e sequenziato su NovaSeq X System con software v1.4 e cella a flusso C8 (63 replicati tecnici). I dati HMW di TruPath Genome sono stati generati con DNA estratto mediante metodi HMW e sequenziato su NovaSeq X System con software v1.4 e cella a flusso C8 (64 replicati tecnici). Dati della chimica PacBio con SPRQ ricavati dal materiale pubblicato.⁶ La precisione misura la percentuale di veri positivi. Il richiamo rappresenta la percentuale di veri negativi rispetto al numero totale di varianti in un campione. Il punteggio F1 (%) è un calcolo dei risultati veri positivi e veri negativi sotto forma di percentuale sui risultati totali. Punteggi più elevati indicano maggiore accuratezza in base ai dati di riferimento.

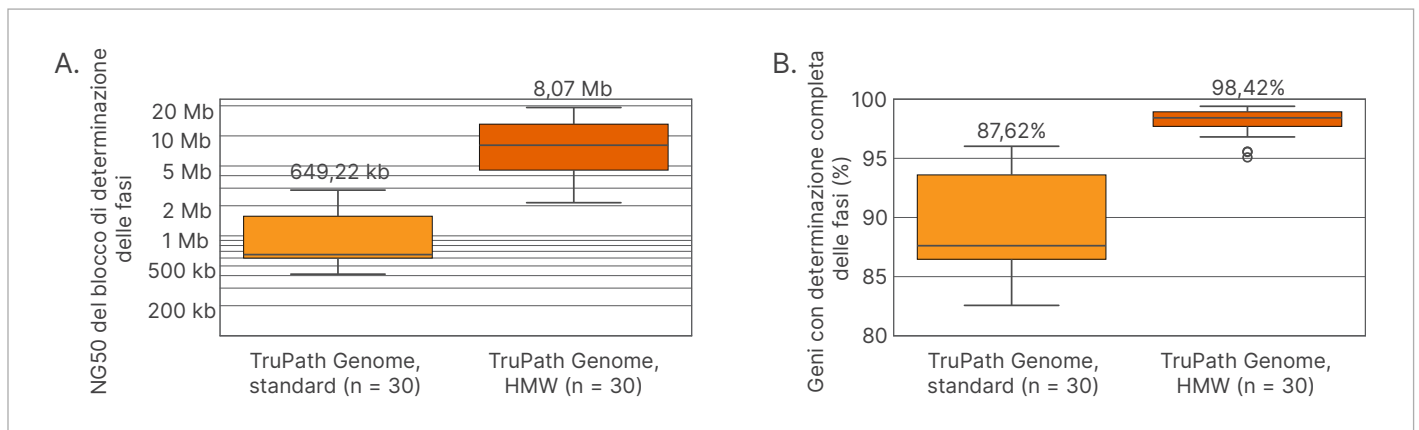


Figura 7: determinazione delle fasi su distanze ultralunghe con TruPath Genome su 30 linee cellulari corrispondenti con estrazione HMW e standard

Il blocco di fase viene misurato sui cromosomi da 1 a 22 utilizzando la determinazione delle fasi DRAGEN in DRAGEN Germline v4.5.2. I dati di TruPath Genome sono stati ottenuti utilizzando 30 campioni Coriell di diverse linee cellulari con estrazioni standard e ad alto peso molecolare (HMW). L'NG50 del blocco di fase corrisponde alla lunghezza del blocco di fase una volta che il 50% della regione target (genoma o altro) è stato sottoposto a determinazione delle fasi. Una tecnologia che non è in grado di determinare le fasi nel 50% di una determinata regione target avrà un NG50 pari a zero coppie di basi. La percentuale di geni con determinazione completa delle fasi corrisponde alla percentuale delle regioni geniche di un determinato elenco di geni (Gencode v44 geni.gtf) interamente contenute in un singolo blocco di fase.

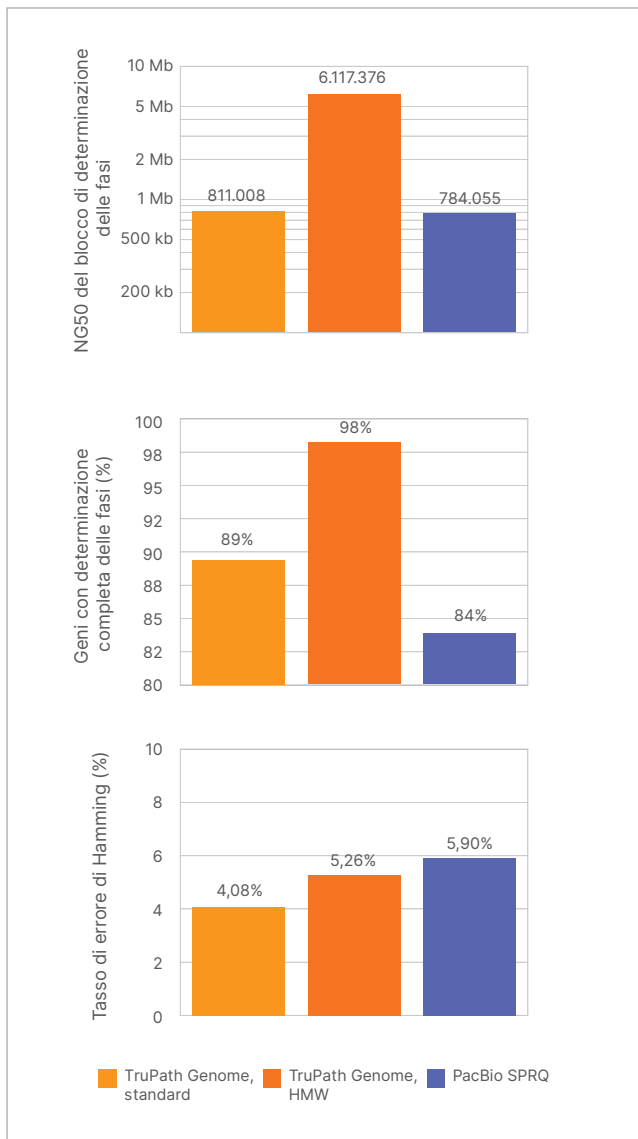


Figura 8: sequenziamento con determinazione delle fasi ottenuto con TruPath Genome, confrontato con i dati del truth set HG002

Il blocco di fase viene misurato sui cromosomi da 1 a 22 utilizzando la determinazione delle fasi DRAGEN in DRAGEN Germline v4.5.2. Il VCF con determinazione delle fasi per i dati HiFi (PacBio SPRQ) è stato ottenuto da https://downloads.pacbcloud.com/public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2/. I dati di TruPath Genome sono stati ottenuti utilizzando replicati di sequenziamento da campioni di linee cellulari HG002 con campioni estratti mediante metodi standard e HMW. L'NG50 del blocco di fase corrisponde alla lunghezza del blocco di fase una volta che il 50% della regione target (genoma o altro) è stato sottoposto a determinazione completa delle fasi. La percentuale di geni con determinazione completa delle fasi corrisponde alla percentuale delle regioni geniche di un determinato elenco di geni (Gencode v44 geni.gtf) interamente contenute in un singolo blocco di fase. Il tasso di errore di Hamming della determinazione delle fasi è stato confrontato con il truth set T2T Q100⁹ scaricato da https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/analysis/NIST_HG002_DraftBenchmark_defrabbV0.020-20250117/GRCh38_HG2-T2TQ100-V1.1_smvar.vcf.gz.

TruPath Genome si è dimostrato in grado di effettuare la determinazione completa delle fasi con una mediana pari a circa il 91% di tutti i geni con estrazione standard del DNA e a circa il 97% dei geni con estrazione ad alto peso molecolare (HMW). Inoltre, TruPath Genome ha determinato le fasi in circa il 98%¹ delle SNV eterozigoti nel DNA estratto sia con metodo standard sia con approccio HMW.

I dati di TruPath Genome con determinazione delle fasi DRAGEN per il DNA estratto con metodo HMW riducono gli errori di Hamming** (circa il 4%) rispetto agli strumenti disponibili al pubblico, come HapCut2 (circa il 6%).¹⁰

Rilevamento delle varianti in regioni ad alta omologia

Il rilevamento accurato delle varianti nelle regioni paraloghe utilizzando letture brevi standard risulta difficile a causa dell'elevata omologia delle sequenze, che comporta ambiguità della mappatura delle letture ed errori di rilevamento delle varianti. Spesso, per risolvere i casi ambigui, sono necessari costosi saggi di conferma, quali amplificazione mediante sonde basate su ligazione multipla (MLPA, Multiplex Ligation-Dependent Probe Amplification) o reazione a catena della polimerasi a lungo raggio (LR-PCR, Long-Range Polymerase Chain Reaction). TruPath Genome combina le informazioni di prossimità della cella a flusso con un nuovo algoritmo di rilevamento congiunto multiregione (MRJD, Multiregion Joint Detection) per consentire l'identificazione affidabile delle varianti, *de novo* e con risoluzione basata su aptotipo, nei geni paraloghi (Figura 9 e Figura 10).

Al fine di ottenere una mappatura altamente precisa mediante MRJD, TruPath Genome mira a undici geni (Tabella 1) noti per avere copie in tandem o non in tandem distribuite in tutto il genoma e genera identificazione di varianti piccole e ricostruzioni con risoluzione basata su aptotipo (Figura 10).

Tabella 1: geni target dell'analisi MRJD con TruPath Genome

<i>CFHR</i>	<i>NCF1</i>	<i>SMN1/2</i>
<i>CYP2D6</i>	<i>PMS2</i>	<i>STRC</i>
<i>CYP11B1</i>	<i>RCCX (CYP21A2 e TNXB)</i>	<i>USP8</i>

¹ La percentuale di varianti eterozigoti con determinazione delle fasi corrisponde alla percentuale di SNV eterozigoti con determinazione delle fasi calcolata dividendo il numero di SNV con determinazione delle fasi per il numero di SNV eterozigoti.

** Gli errori di Hamming si verificano quando le varianti vengono assegnate al filamento materno o paterno errato.

Risoluzione migliorata per le brevi ripetizioni in tandem (STR)

Le brevi ripetizioni in tandem (STR, Short Tandem Repeat) sono regioni ripetitive del genoma che possono raggiungere lunghezze superiori all'intervallo normale e causare mutazioni associate a molte malattie genetiche, tra cui la sindrome dell'X fragile, la sclerosi laterale amiotrofica e la malattia di Huntington.¹¹ La stima accurata delle dimensioni e il recupero delle letture di queste regioni risultano difficoltosi, perché gli alleli espansi più rilevanti spesso superano la lunghezza di lettura dei dati di sequenziamento standard a lettura breve. Grazie alla combinazione di informazioni di prossimità e dati a lettura breve di alta qualità, TruPath Genome migliora il recupero delle letture necessarie per l'accurato dimensionamento dell'STR, nonché la risoluzione mediante la genotipizzazione delle due copie utilizzando le informazioni con determinazione delle fasi (Figura 11).

Informazioni su distanze ultralunghe per le SV complesse

La capacità di TruPath Genome di risolvere i riarrangiamenti strutturali di grandi dimensioni è il frutto, tra le altre cose, di nuove funzionalità che vanno oltre i tradizionali confronti delle prestazioni di identificazione di varianti. TruPath Genome acquisisce le informazioni sulle letture dai cluster prossimali^{††} tra qualsiasi coppia di regioni del genoma, consentendo di creare rappresentazioni visive ad alta risoluzione delle mappe strutturali del genoma, i cosiddetti "grafici di colocalizzazione" (Figura 12).

†† I cluster prossimali sono cluster fisicamente vicini l'uno all'altro sulla cella a flusso.

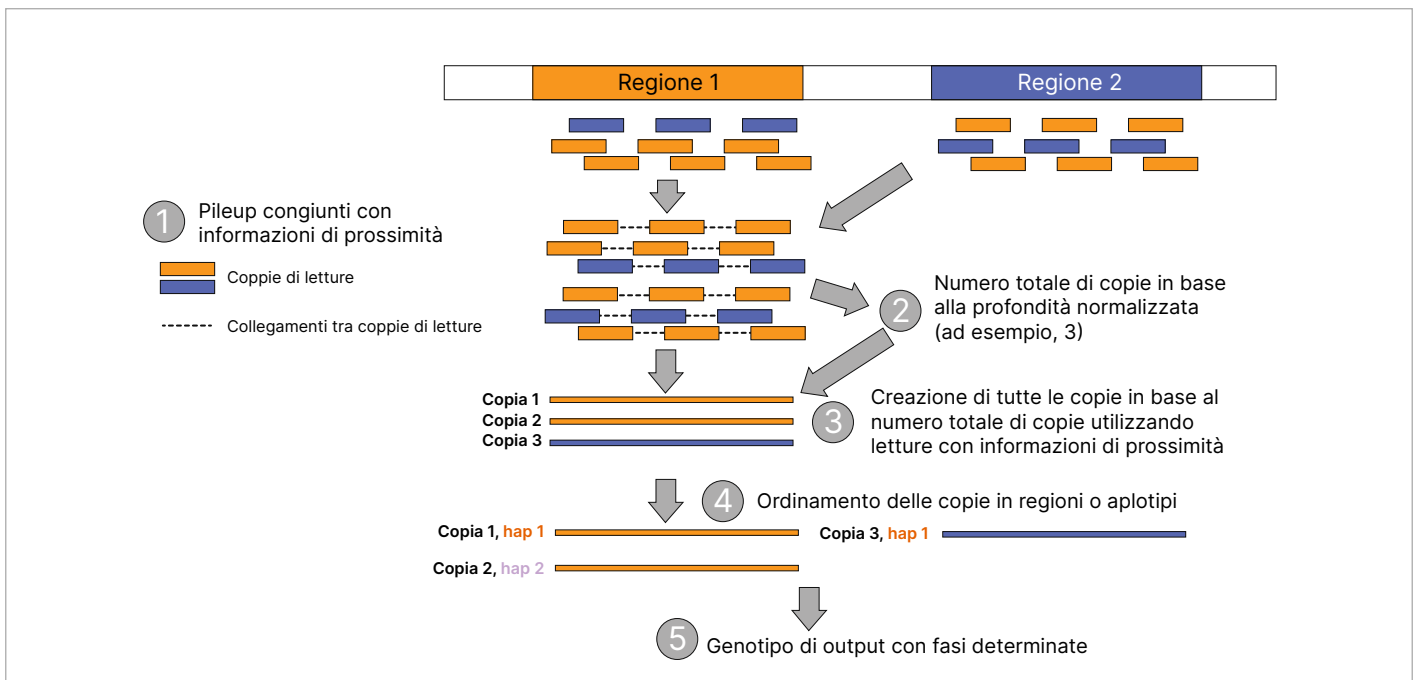
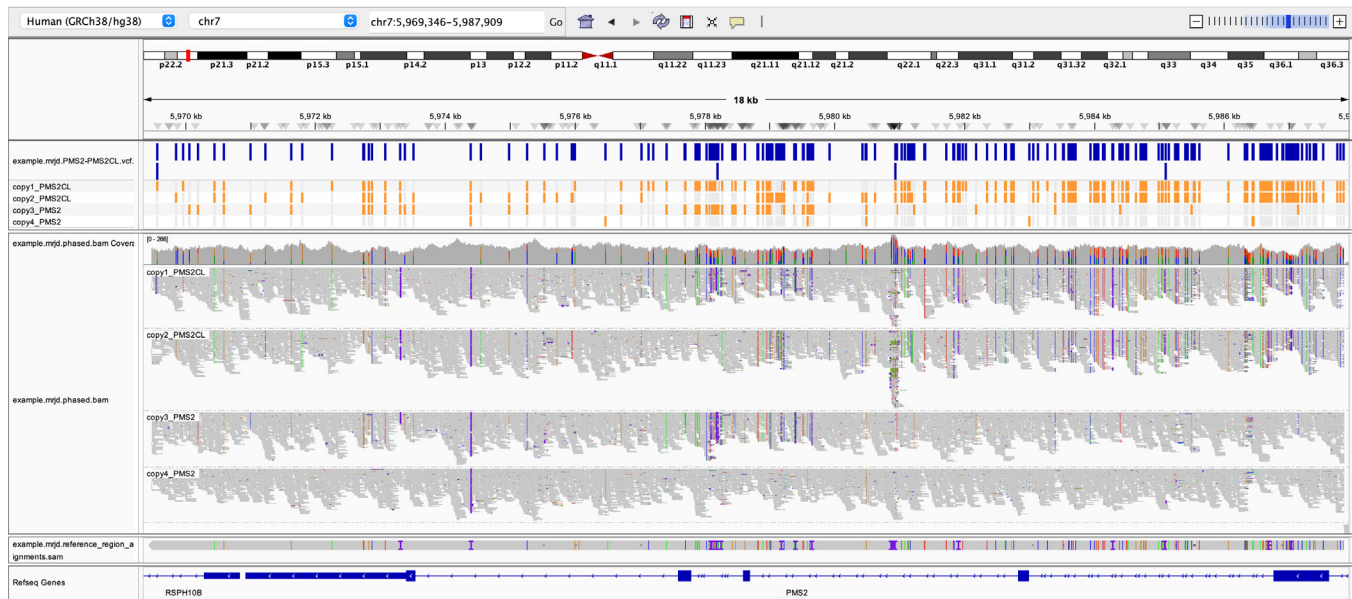


Figura 9: schema del metodo MRJD utilizzato per rilevare le varianti nelle regioni di duplicazione segmentale

1. Per un determinato set di regioni duplicate segmentalmente, l'MRJD estrae tutte le letture mappate in una qualsiasi delle copie nel set di regioni per l'analisi congiunta. Vengono impiegate le informazioni di prossimità per stabilire i collegamenti tra tali letture e identificare i set di letture derivanti dalla stessa molecola di partenza. 2. Il conteggio di tutte le letture estratte nella fase 1 viene quindi normalizzato, corretto per il contenuto in GC e utilizzato per stimare il numero totale di copie della sequenza duplicata segmentalmente nel campione di input. 3. Viene utilizzato un approccio di genotipizzazione congiunta basato sulle informazioni relative al numero di copie della regione stimato nella fase 2 e applicato all'intero set di letture con collegamenti di prossimità allo scopo di identificare e sottoporre a determinazione delle fasi tutte le varianti in ciascuna copia della sequenza duplicata segmentalmente nel campione di input. 4. Ogni copia con determinazione completa delle fasi viene assegnata a una specifica posizione nel genoma di riferimento (quando le regioni si trovano in posizioni distali nel riferimento) oppure viene sottoposta a determinazione delle fasi insieme alle altre copie per definire gli aplotipi (possibile solo quando le regioni di duplicazione segmentale sono in tandem o vicine nel riferimento). 5. Le varianti con determinazione completa delle fasi per tutte le copie, l'assegnazione dell'aplotipo di ciascuna copia, l'assegnazione delle copie a ciascuna posizione del genoma di riferimento e le letture assegnate a ciascuna copia sono estratte nei file VCF, JSON e BAM specifici per l'MRJD.

A. Gene *PMS2*



B. Gene *SMN*

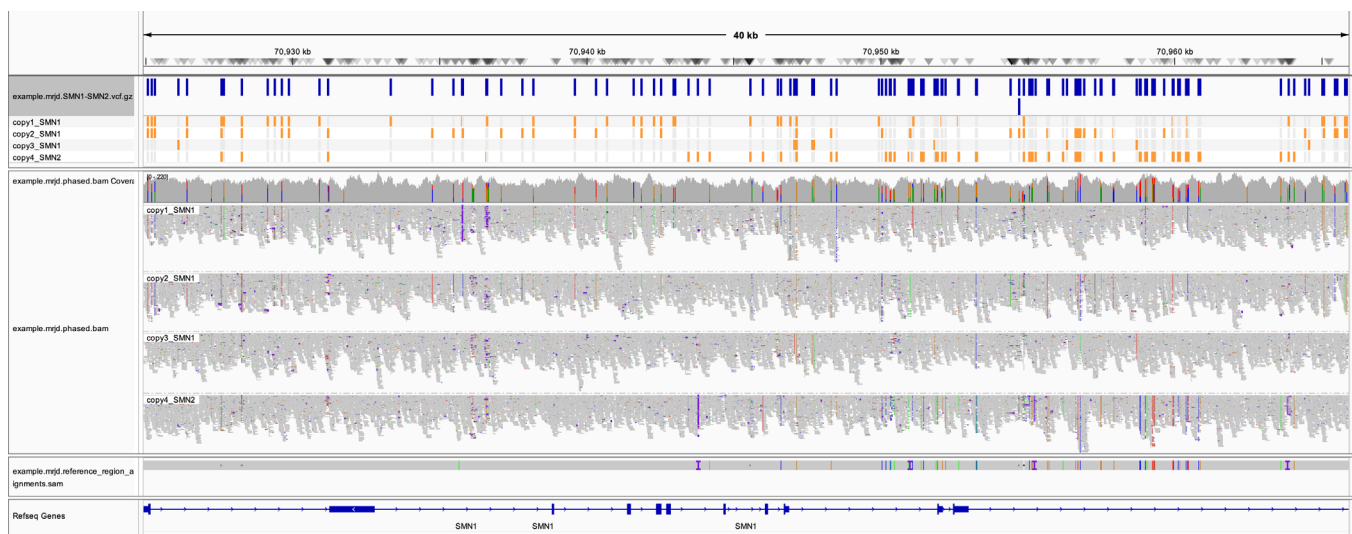


Figura 10: esempio di output delle identificazioni di varianti e dell'assegnazione delle letture nei geni paraloghi con TruPath Genome ed MRJD

(A) Identificazioni di varianti con determinazione delle fasi e letture di supporto associate per tutte le copie della coppia gene-pseudogene *PMS2/PMS2CL* rappresentata nella posizione *PMS2* del genoma di riferimento. Delle quattro copie, due sono state identificate come copie di *PMS2CL* (pseudogene) e due come copie di *PMS2* (gene attivo) sulla base dei collegamenti di prossimità alle regioni fiancheggianti *PMS2/PMS2CL*. Copy3 (*PMS2*) contiene un possibile evento di conversione genica che si sovrappone agli esoni 13 e 14, indicato dalla presenza di varianti corrispondenti alla sequenza di riferimento *PMS2CL*. L'allineamento tra la sequenza di riferimento *PMS2CL* e la sequenza di riferimento *PMS2* è rappresentato nel terzo riquadro (dall'alto verso il basso). (B) Identificazioni di varianti con determinazione delle fasi e letture di supporto associate per tutte le copie della coppia gene-pseudogene *SMN1/SMN2* rappresentata nella posizione *SMN1* del genoma di riferimento. Delle quattro copie, tre sono state identificate come *SMN1* (gene attivo) e due come *SMN2* (gene inattivo) secondo la base nella posizione *SMN1* c.840 in ciascuna delle copie sottoposte a determinazione delle fasi (le copie *SMN1* hanno la base C e le copie *SMN2* hanno la base T in detta posizione).

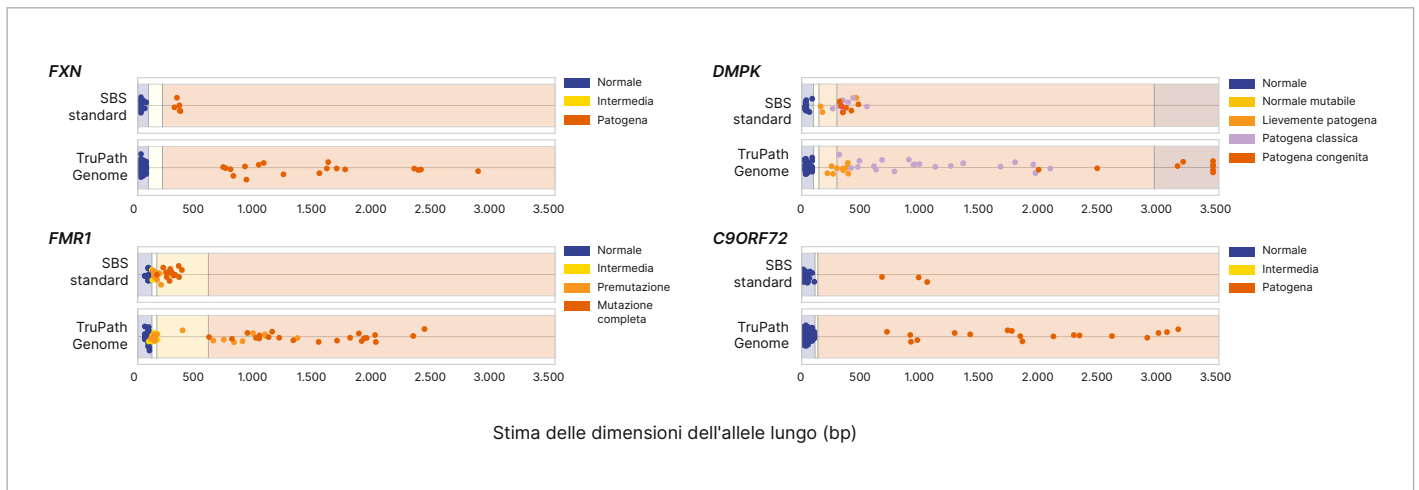


Figura 11: con TruPath Genome è possibile ottenere misurazioni più accurate delle espansioni STR

Le classificazioni corrette delle STR sono rappresentate dal colore di ciascun punto. Le strisce colorate rappresentano i diversi intervalli di classificazione utilizzati per definire le diverse classi delle espansioni in ciascun locus. L'asse x indica le dimensioni delle STR stimate da DRAGEN STR con TruPath Genome o l'SBS standard (Illumina DNA PCR-Free Prep) come input. Le classificazioni di TruPath Genome risultano notevolmente più coerenti con le classificazioni corrette e coprono un più ampio intervallo di lunghezze delle STR. Replicati totali per SBS standard: 78 campioni biologici (78 individui univoci). Replicati totali per TruPath Genome: 135 campioni biologici (42 individui univoci).

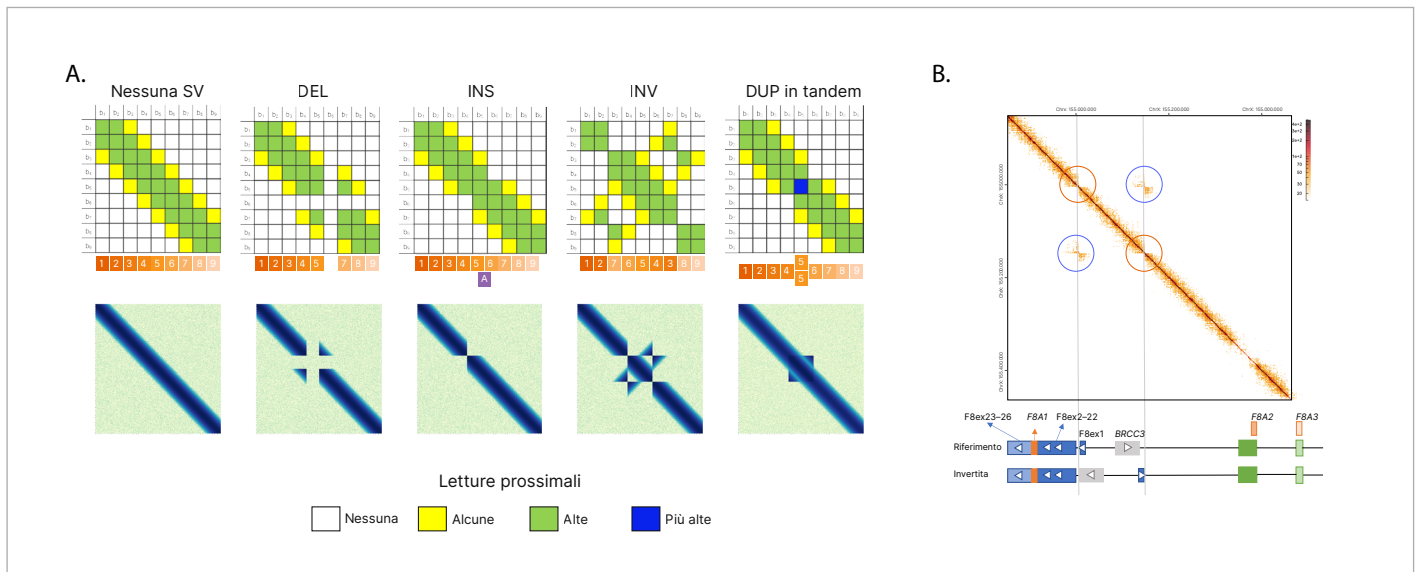


Figura 12: visualizzazione degli eventi di varianti strutturali mediante i grafici di colocalizzazione di TruPath Genome

(A) I conteggi delle letture prossimali (spazio della cella a flusso) mappate su tutte le coppie di intervalli genomici vengono valutati per generare una mappa di colocalizzazione bidimensionale. Gli assi X e Y rappresentano gli intervalli del genoma di riferimento. Un elevato numero di connessioni tra due intervalli indica che questi ultimi sono vicini l'uno all'altro nello spazio genomico nel campione di input. In una regione senza varianti strutturali, gli intervalli genomici che si trovano vicini nel genoma di riferimento sono vicini anche nel campione di input e quindi appaiono come numero elevato di connessioni vicino alla diagonale del grafico di colocalizzazione. Quando è presente una variante strutturale, gli intervalli genomici che si trovano vicini nel genoma di riferimento potrebbero non essere più vicini nel genoma del campione di input e gli intervalli genomici che sono lontani nel riferimento potrebbero essere prossimali nel campione di input. Tali variazioni nella connettività si riflettono sulle mappe di colocalizzazione come segnali fuori diagonale di forme diverse, nonché come mancanza di connettività lungo la diagonale. (B) Esempio di mappa di colocalizzazione per un'inversione genica dell'introne 1 nel locus del gene *F8*, associata a grave emofilia A. Si notino il segnale previsto a forma di clessidra nella regione fuori diagonale (cerchi blu) e la mancanza di segnale lungo la diagonale (cerchi arancioni) ai confini dell'evento. Sotto il grafico è riportata una rappresentazione schematica del locus nel genoma di riferimento e della regione genica invertita.

Il più semplice flusso di lavoro per il sequenziamento dell'intero genoma

Illumina TruPath Genome, basato sulla fondamentale tecnologia di lettura con mappatura di prossimità, trasforma il sequenziamento del genoma umano. L'innovativo flusso di lavoro elimina la tradizionale preparazione delle librerie per una semplicità senza precedenti, mentre i dati di prossimità preservano le informazioni sui filamenti di DNA su lunghe distanze sfruttando dati di sequenziamento standard a lettura breve altamente accurati. Il risultato è un saggio che combina i punti di forza delle letture brevi in quanto a qualità dei dati e il rilevamento altamente accurato di varianti piccole con una migliore copertura della mappatura del genoma umano e un potente strumento per lo studio delle malattie genetiche.

Maggiori informazioni →

[Illumina TruPath Genome](#)

[Tecnologia di lettura con mappatura di prossimità](#)

[NovaSeq X Series](#)

Informazioni per gli ordini	
Prodotto	N. di catalogo
Illumina TruPath Genome (NovaSeq X C8 flow cell)	20157405
Illumina TruPath Genome (NovaSeq X C2 flow cell)	20157406

Bibliografia

1. Ebbert MTW, Jensen TD, Jansen-West K, et al. [Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight](#). *Genome Biol.* 2019;20(1):97. Pubblicato il 20 maggio 2019. doi:10.1186/s13059-019-1707-2
2. Ryan NM, Corvin A. [Investigating the dark-side of the genome: a barrier to human disease variant discovery?](#) *Biol Res.* 2023;56(1):42. Pubblicato il 20 luglio 2023. doi:10.1186/s40659-023-00455-0
3. Illumina. TruPath Genome performance with samples of varying type and quality. [illumina.com/content/dam/illumina-marketing/documents/products/technotes/trupath-genome-sample-info-tech-note-m-gl-03932](https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/trupath-genome-sample-info-tech-note-m-gl-03932). Pubblicato a febbraio 2026. Consultato in data XX febbraio 2026.
4. Wadsworth ME, Page ML, Aguzzoli Heberle B, Miller JB, Steely C, Ebbert MTW. [Sequencing the gaps: dark genomic regions persist in CHM13 despite long-read advances](#). Prestampa. *bioRxiv.* 2025;2025.05.23.655776. Pubblicato il 28 maggio 2025. doi:10.1101/2025.05.23.655776
5. Ultima Genomics. Genome-in-a-Bottle (GIAB) samples HG001-HG005. <https://cdn.sanity.io/files/l7780ks7/production-2024/0a1b6a62a6da3e3fcafb81cad4c8ff2ffe85dd41.pdf>. Pubblicato a febbraio 2025. Consultato il 29 gennaio 2026.
6. PacBio. Indice di [/public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2](https://downloads.pacbcloud.com/public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2/). https://downloads.pacbcloud.com/public/revio/2024Q4/WGS/GIAB_trio/HG002_rep1/analysis/v3.0.2/. Consultato il 29 gennaio 2026.
7. Asri M, Chang PC, Mier JC, et al. [Pangenome-aware DeepVariant](#). Prestampa. *bioRxiv.* 2025;2025.06.05.657102. Pubblicato il 6 giugno 2025. doi:10.1101/2025.06.05.657102
8. Browning SR, Browning BL. [Haplotype phasing: existing methods and new developments](#). *Nat Rev Genet.* 2011; 12(10):703-714. doi: 10.1038/nrg3054
9. Hansen NF, Dwarshuis N, Ji HJ, et al. [A complete diploid human genome benchmark for personalized genomics](#). Prestampa. *bioRxiv.* 2025;2025.09.21.677443. Pubblicato il 21 settembre 2025. doi:10.1101/2025.09.21.677443
10. Dati in archivio. Illumina, Inc., 2026.
11. Steely CJ, Watkins WS, Baird L, Jorde LB. [The mutational dynamics of short tandem repeats in large, multigenerational families](#). *Genome Biol.* 2022;23(1):253. Pubblicato il 12 dicembre 2022. doi:10.1186/s13059-022-02818-4



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

© 2026 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-03931 ITA v1.0