

Strumento NextSeq™ 550Dx

PER USO DIAGNOSTICO IN VITRO
SOLO PER L'ESTERO

N. di catalogo 20005715

Uso previsto

Lo strumento NextSeq 550Dx Illumina è destinato al sequenziamento di librerie di DNA quando utilizzato con saggi per diagnostica *in vitro*. Lo strumento NextSeq 550Dx deve essere utilizzato con determinati reagenti per diagnostica *in vitro* e software analitici registrati, autorizzati o approvati.

Principi della procedura

Lo strumento NextSeq 550Dx Illumina è progettato per il sequenziamento di librerie di DNA con saggi per diagnostica *in vitro* e la sua destinazione è l'uso da parte di personale di laboratorio clinico qualificato e addestrato all'uso di procedure diagnostiche *in vitro* eseguite in un laboratorio clinico. Come input, NextSeq 550Dx utilizza librerie generate da DNA in cui gli indici campioni e le sequenze di cattura vengono aggiunti ai target amplificati. Le librerie di campioni sono catturate su una cella a flusso e sequenziate sullo strumento utilizzando la chimica di sequenziamento mediante sintesi (SBS, Sequencing By Synthesis). La chimica SBS utilizza un metodo che fa uso di terminatori reversibili per rilevare le singole basi nucleotidiche marcate con colorante fluorescente man mano che vengono incorporate in filamenti di DNA in estensione. Il software Real-Time Analysis (RTA) esegue l'analisi delle immagini e l'identificazione delle basi e assegna un punteggio qualitativo a ciascuna base per ciascun ciclo di sequenziamento. Al termine dell'analisi primaria, è possibile eseguire l'analisi secondaria sullo strumento per elaborare le identificazioni delle basi. NextSeq 550Dx utilizza diversi moduli per l'analisi secondaria in base al flusso di lavoro. Per i moduli Germline Variant e Somatic Variant, l'elaborazione include demultiplex, generazione di file FASTQ, allineamento, identificazione delle varianti e generazione di file nel formato di identificazione delle varianti (Variant Call Format, VCF). I file VCF contengono informazioni sulle varianti trovate in posizioni specifiche in un genoma di riferimento.

Configurazione dual boot

NextSeq 550Dx include una configurazione dual boot per consentire l'uso dello strumento sia in modalità diagnostica (Diagnostic, Dx) sia in modalità Solo a uso di ricerca (Research Use Only, RUO). I saggi per il sequenziamento per la diagnostica *in vitro*, inclusi i moduli Germline Variant e Somatic Variant, sono eseguiti in modalità diagnostica. I reagenti per il sequenziamento per la diagnostica in-vitro (In-Vitro Diagnostic, IVD) possono essere utilizzati solo in modalità diagnostica. Le caratteristiche delle prestazioni e i limiti della procedura per lo strumento NextSeq 550Dx sono stati stabiliti utilizzando i moduli Germline Variant e Somatic Variant in modalità diagnostica.

Limiti della procedura

- 1 Per uso diagnostico *in vitro*.
- 2 I moduli Germline Variant e Somatic Variant, quando utilizzati con NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) o con NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles), sono in grado di fornire:
 - ▶ Output di sequenziamento ≥ 90 gigabasi (Gb)
 - ▶ Lunghezza di lettura (in corse paired-end) da 2 x 150 coppie di basi (bp)
 - ▶ Basi con un punteggio Q30 pari o superiore a $\geq 75\%$ alla lunghezza di lettura di 2 x 150 bp

- Un numero pari o maggiore al 75% di basi presenta un punteggio qualitativo su scala Phred di ≥ 30 , indicando che l'accuratezza dell'identificazione delle basi supera il 99,9%
- 3 Letture con Indel (inserzioni, delezioni o combinazioni) la cui lunghezza di lettura è superiore a 25 bp non vengono allineate dal software del saggio. Di conseguenza, le indel di lunghezza superiore a 25 bp non sono rilevabili dal software del saggio.
 - 4 Il software del saggio potrebbe non allineare le letture degli ampliconi con contenuto estremo della variante, risultando in regioni riportate come wild type. Tale contenuto estremo include:
 - ▶ Letture contenenti più di tre indel
 - ▶ Letture con lunghezza di almeno 30 bp e con contenuto di variante di singolo nucleotide (Single Nucleotide Variant, SNV) superiore a 4% della lunghezza totale dell'amplicone target (escluse le regioni della sonda)
 - ▶ Letture di lunghezza inferiore a 30 bp con contenuto SNV superiore al 10% della lunghezza totale dell'amplicone (incluse le regioni della sonda)
 - 5 Le varianti ampie, incluse le varianti di più nucleotidi (MNV, Multi-Nucleotide Variant) e ampie indel, potrebbero essere riportate nel file di output VCF come varianti separate più piccole.
 - 6 Le varianti delle delezioni possono essere filtrate o non individuate quando si estendono su due ampliconi rilevati sulle tile se la lunghezza della delezione è maggiore o uguale alla sovrapposizione tra gli ampliconi rilevati sulle tile.
 - 7 Il sistema non può rilevare le indel, se si verificano direttamente accanto a un primer e dove non è presente un amplicone sovrapposto. Per le regioni con ampliconi sovrapposti, il saggio non può rilevare le delezioni quando la regione di sovrapposizione è inferiore alla dimensione della delezione da rilevare. Ad esempio, se la regione di sovrapposizione tra due ampliconi adiacenti è di due basi, il saggio non può rilevare alcuna delezione includendo entrambe quelle basi. Può essere rilevata una delezione di una singola base a una qualsiasi di quelle basi.
 - 8 Come per qualsiasi flusso di lavoro di preparazione delle librerie basato su ibridazione, i polimorfismi, le mutazioni, le inserzioni o le delezioni latenti nelle regioni che legano gli oligonucleotidi possono incidere sugli alleli sondati e, di conseguenza, sulle identificazioni effettuate durante il sequenziamento. Ad esempio:
 - ▶ Una variante nella fase con una variante nella regione del primer potrebbe non essere amplificata e fornire un falso negativo.
 - ▶ Le varianti nella regione del primer potrebbero impedire l'amplificazione dell'allele di riferimento e fornire un'identificazione della variante omozigote errata.
 - ▶ Le varianti delle indel nella regione del primer potrebbero fornire un'identificazione falso positiva al termine della lettura adiacente al primer.
 - 9 Le indel possono essere filtrate a causa di distorsioni del filamento se si verificano accanto alla fine di una lettura e sono sottoposte a soft-clipping durante l'allineamento.
 - 10 Non sono state convalidate piccole MNV e vengono solo riportate nel modulo Somatic Variant.
 - 11 Le delezioni sono riportate nel file VCF alla coordinata della base precedente in base al formato VCF. Pertanto, prendere in considerazione le varianti adiacenti prima di riportare una singola identificazione delle basi come un riferimento omozigote.
 - 12 Limitazioni specifiche per il modulo Germline Variant:
 - ▶ Lo strumento NextSeq 550Dx, utilizzando il modulo Germline Variant di Local Run Manager per NextSeq 550Dx, è progettato per fornire risultati qualitativi per l'identificazione delle varianti della linea germinale (ad es., omozigoti, eterozigoti, wild type).
 - ▶ Quando utilizzato con il modulo Germline Variant, la copertura minima per l'amplicone necessaria per l'accurata identificazione delle varianti è di 150x. Di conseguenza, sono richiesti frammenti di DNA in grado di supportare una copertura di 150; questo valore equivale a 300 letture paired-end sovrapposte. Il numero di campioni e il numero totale di basi target incidono sulla copertura. Il contenuto in GC e altro contenuto genomico possono incidere sulla copertura.
 - ▶ La variazione del numero di copie può incidere sulla possibilità che una variante venga identificata come omozigote o eterozigote.

- ▶ Le varianti in un determinato contesto ripetitivo sono filtrate nei file VCF. Il filtro RMxN per le ripetizioni viene utilizzato per filtrare le varianti se tutta o parte della sequenza della variante è presente ripetutamente nel genoma di riferimento adiacente alla posizione della variante. Per l'identificazione delle varianti della linea geminale, sono richieste almeno nove ripetizioni nel riferimento affinché una variante venga filtrata. Vengono prese in considerazione solo le ripetizioni con una lunghezza fino a 5 bp (R5x9).
 - ▶ Una indel e una SNV su un singolo locus possono risultare in una sola variante riportata.
- 13 Limitazioni specifiche per il modulo Somatic Variant.
- ▶ Lo strumento NextSeq 550Dx, utilizzando il modulo Somatic Variant di Local Run Manager per NextSeq 550Dx, è progettato per offrire risultati qualitativi per l'identificazione delle varianti somatiche (ad es., presenza di una variante somatica con una frequenza della variante maggiore o uguale a 0,026 con un limite del rilevamento di 0,05).
 - ▶ Quando utilizzato con il modulo Somatic Variant, la copertura minima per l'amplicone necessaria per l'accurata identificazione delle varianti è di 450x per il pool di oligonucleotidi. Di conseguenza, per il pool di oligonucleotidi sono richiesti frammenti di DNA in grado di supportare una copertura di 450; questo valore equivale a 900 letture paired-end sovrapposte. Il numero di campioni e il numero totale di basi target incidono sulla copertura. Il contenuto in GC e altro contenuto genomico possono incidere sulla copertura.
 - ▶ Per l'identificazione delle varianti somatiche, sono richieste almeno sei ripetizioni nel riferimento affinché una variante venga filtrata e vengono prese in considerazione solo le ripetizioni con una lunghezza fino a 3 bp (R3x6).
 - ▶ Il modulo Somatic Variant non è in grado di differenziare tra le varianti delle linee geminali e le varianti somatiche. Il modulo è progettato per rilevare le varianti su un intervallo di frequenze delle varianti, ma la frequenza della variante non può essere utilizzata per differenziare le varianti somatiche dalle varianti della linea geminale.
 - ▶ I tessuti normali contenuti nel campione incidono sul rilevamento delle varianti. Il limite del rilevamento riportato si basa su una frequenza della variante relativa al DNA totale estratto sia da tessuto di tumore che da tessuto normale.

Componenti del prodotto

- 1 Strumento NextSeq 550Dx (n. di catalogo 20005715)
- 2 I componenti software per lo strumento NextSeq 550Dx includono:

Applicazione software	Funzione	Descrizione
NextSeq 550Dx Operating Software (NOS)	Controlla il funzionamento dello strumento	L'applicazione software NOS gestisce il funzionamento dello strumento durante il sequenziamento e genera immagini da utilizzare con il software Real-Time Analysis (RTA).
Real-time Analysis Software (RTA)	Esegue l'analisi primaria	L'applicazione software RTA converte le immagini generate da NOS per ciascuna tile per ciclo della corsa di sequenziamento in file di identificazione delle basi che rappresentano gli input per i moduli di analisi di Local Run Manager. L'applicazione software RTA non contiene un'interfaccia utente.
Local Run Manager	L'interfaccia per la selezione del modulo	Il software Local Run Manager è una soluzione integrata sullo strumento per la gestione degli utenti, la selezione del modulo di analisi appropriato e il monitoraggio dello stato.

Applicazione software	Funzione	Descrizione
Modulo Somatic Variant	Esegue l'analisi secondaria	Questo modulo di analisi di Local Run Manager elabora le identificazioni delle basi mediante l'analisi secondaria. L'elaborazione include demultiplex, generazione di file FASTQ, allineamento, identificazione delle varianti e creazione di report. L'identificatore delle varianti (Pisces) genera file VCF che contengono le informazioni relative alle varianti individuate in determinate posizioni in un genoma di riferimento e includono la frequenza della variante misurata.
Modulo Germline Variant	Esegue l'analisi secondaria	Questo modulo di analisi di Local Run Manager elabora le identificazioni delle basi mediante l'analisi secondaria. L'elaborazione include demultiplex, generazione di file FASTQ, allineamento, identificazione delle varianti e creazione di report. L'identificatore delle varianti (Pisces) genera file VCF che contengono le informazioni relative alle varianti individuate in determinate posizioni in un genoma di riferimento e identifica ciascuna variante come eterozigote od omozigote.

Condizioni di funzionamento

Elemento	Specifica
Temperatura	Mantenere nel laboratorio una temperatura compresa tra 19 °C e 25 °C (22 °C ± 3 °C). Questa è la temperatura operativa dello strumento. Durante una corsa, evitare che la temperatura ambiente subisca sbalzi superiori a ± 2 °C.
Umidità	Mantenere l'umidità relativa, senza condensa, nell'intervallo 20-80%.

Apparecchiature e materiali

Apparecchiature e materiali di consumo richiesti, venduti separatamente

NextSeq 550Dx High Output Reagent Kit v2.5 (75 cycles), N. di catalogo 20028870

NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles), N. di catalogo 20028871

Apparecchiature e materiali richiesti, non forniti

Materiali di consumo forniti dall'utente per le corse di sequenziamento

Materiale di consumo	Fornitore	Scopo
Salviettine imbevute di alcol isopropilico al 70% oppure etanolo al 70%	VWR, n. di catalogo 95041-714 (o equivalente) Fornitore di laboratorio generico	Pulizia della cella a flusso e per uso generico
Panno da laboratorio a bassissimo rilascio di particelle	VWR, n. di catalogo 21905-026, o equivalente	Pulizia della cella a flusso e per uso generico

Materiali di consumo forniti dall'utente per la manutenzione dello strumento

Materiale di consumo	Fornitore	Scopo
NaOCl, 5% (ipoclorito di sodio)	Sigma-Aldrich, n. di catalogo 239305 (o equivalente da laboratorio)	Lavaggio dello strumento utilizzando un lavaggio post-corsa manuale; diluito allo 0,12%
Tween 20	Sigma-Aldrich, n. di catalogo P7949	Lavaggio dello strumento utilizzando le opzioni di lavaggio manuale, diluito allo 0,05%

Materiale di consumo	Fornitore	Scopo
Acqua da laboratorio	Fornitore di laboratorio generico	Lavaggio dello strumento (lavaggio manuale)
Filtro dell'aria	llumina, n. di catalogo 20022240	Pulizia dell'aria aspirata dallo strumento per il raffreddamento

Linee guida per l'acqua da laboratorio

Per eseguire le procedure dello strumento utilizzare sempre acqua da laboratorio o acqua deionizzata. Non usare mai acqua di rubinetto. Utilizzare solo acqua da laboratorio o gli equivalenti seguenti:

- ▶ Acqua deionizzata
- ▶ PW1 Illumina
- ▶ Acqua con resistività pari a 18 megaohm (MΩ)
- ▶ Acqua Milli-Q
- ▶ Acqua Super-Q
- ▶ Acqua sterile per biologia molecolare

Avvertenze e precauzioni

ATTENZIONE La legge federale stabilisce che la vendita, l'uso o l'indicazione all'uso di questo dispositivo deve avvenire da parte o dietro prescrizione di un medico autorizzato dalla legge dello stato in cui esercita.

- 1 **Alcuni componenti dei reagenti forniti da Illumina da utilizzare con lo strumento NextSeq 550Dx contengono composti chimici potenzialmente pericolosi. L'inalazione, l'ingestione, il contatto con la pelle o con gli occhi possono causare lesioni personali. Indossare l'attrezzatura protettiva, inclusi protezione per gli occhi, guanti e indumento da laboratorio appropriato per evitare i rischi di esposizione. Manipolare i reagenti usati come rifiuti chimici e smaltirli in base alle leggi e alle regolamentazioni applicabili a livello regionale, nazionale e locale.** Per maggiori informazioni sulle considerazioni ambientali, sulla sicurezza e sulla salute, vedere le schede di sicurezza (Safety Data Sheet, SDS) alla pagina Web support.illumina.com/sds.html.
- 2 Riferire immediatamente qualsiasi incidente serio relativo a questo prodotto a Illumina e alle autorità competenti degli stati membri nei quali l'utente e il paziente sono residenti.
- 3 Manipolare tutti i campioni di sangue come se fossero contagiosi per il virus dell'immunodeficienza umana (Human Immunodeficiency Virus, HIV), il virus dell'epatite B (Human Hepatitis B Virus, HBV) o altri agenti patogeni a trasmissione ematica (precauzioni universali).
- 4 In caso contrario le procedure indicate potrebbero fornire risultati errati o una significativa riduzione nella qualità del campione.
- 5 Adottare le normali precauzioni di laboratorio. Non pipettare con la bocca. Non mangiare, bere o fumare nelle aree designate per il lavoro. Manipolare i campioni e i reagenti del kit indossando guanti e indumenti da laboratorio monouso. Dopo aver maneggiato i campioni e i reagenti del kit lavarsi bene le mani.
- 6 È necessario adottare pratiche di laboratorio e igiene di laboratorio idonee per impedire la contaminazione di reagenti, strumenti e campioni di DNA genomico con i prodotti della PCR. La contaminazione da PCR può produrre risultati inesatti e inaffidabili.
- 7 Al fine di prevenire la contaminazione, accertarsi che le aree di pre-amplificazione e di post-amplificazione siano dotate di apparecchiatura dedicata (ad es., pipette, punte per pipette, agitatori e centrifughe).
- 8 L'indice per l'accoppiamento dei campioni deve corrispondere esattamente al layout stampato della piastra. Local Run Manager popola automaticamente gli index primer associati con i nomi dei campioni, quando

vengono immessi nel modulo. Si consiglia di verificare gli index primer associati con i campioni prima di avviare la corsa di sequenziamento. La mancata corrispondenza tra il campione e il layout della piastra risulterà in campioni non identificati correttamente e report con risultati errati.

- 9 Si consiglia vivamente l'installazione di un software antivirus, fornito dall'utente, per proteggere il computer dai virus. Consultare il Manuale d'uso per le istruzioni relative all'installazione.
- 10 Non utilizzare NextSeq 550Dx se un qualsiasi pannello è rimosso. Il funzionamento dello strumento con un qualsiasi pannello rimosso crea esposizione potenziale a tensioni di rete e tensioni c.c..
- 11 Non toccare il piano portacelle nello scomparto della cella a flusso. Il riscaldatore in questo scomparto funziona a una temperatura compresa tra 22 °C e 95 °C e potrebbe provocare bruciature.
- 12 Lo strumento pesa circa 84 kg ed eventuali cadute o errata manipolazione possono causare lesioni gravi.

Istruzioni per l'uso

Le seguenti istruzioni per l'uso consentono di utilizzare i moduli Germline Variant e Somatic Variant in modalità diagnostica sullo strumento NextSeq 550Dx utilizzando NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) oppure NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).

Immissione delle informazioni per la corsa

Per istruzioni dettagliate, vedere NextSeq 550Dx Instrument Reference Guide (Guida di consultazioni dello strumento NextSeq 550Dx) (documento n. 1000000009513) e la guida per il modulo Local Run Manager pertinente.

Impostazione dei parametri

- 1 Accedere a Local Run Manager.
- 2 Fare clic su **Create Run** (Crea corsa) e selezionare **Somatic Variant** o **Germline Variant**.
- 3 Immettere un nome che identifichi la corsa dal sequenziamento fino all'analisi.
Utilizzare caratteri alfanumerici, spazi, trattini bassi o trattini.
- 4 **[Facoltativo]** Immettere una descrizione per identificare la corsa.
Utilizzare caratteri alfanumerici, spazi, trattini bassi o trattini.
- 5 Selezionare il numero di campioni e set di indici dall'elenco a discesa.
Prendere in considerazione le seguenti informazioni quando si effettua una selezione.
 - ▶ L'elenco a discesa contiene i numeri dei campioni con un set di indici. Ad esempio, 24-Set 1 indica 24 campioni da analizzare, con indici appartenenti al set di indici 1.
 - ▶ I numeri dei set di indici si riferiscono a diversi set di coppie di indici i5 e i7. Set 1 e Set 2 forniscono entrambi la diversità d'indice. I due set di indici forniti contribuiscono a impedire la deplezione di un singolo set.
 - ▶ Scegliere il numero di campioni che più si avvicina al numero di campioni da analizzare. Se il numero esatto di campioni non è presente nell'elenco, selezionare il numero che più vi si avvicina, ma che sia inferiore al numero di campioni da analizzare. Ad esempio, se si desidera analizzare 18 campioni, selezionare 16 campioni.
 - ▶ I pozzetti dei campioni suggeriti e le combinazioni d'indice che soddisfano i requisiti di diversità dell'indice sono evidenziati in verde.

Importazione di file manifest per la corsa

- 1 Assicurarsi che i file manifest da importare siano disponibili in una posizione di rete accessibile o su un dispositivo USB.
- 2 Selezionare **Import Manifests** (Importa file manifest).
- 3 Individuare il file manifest e selezionare i file manifest da aggiungere.


NOTA Per far sì che i file manifest siano disponibili per tutte le corse utilizzando il modulo di analisi Germline Variant o Somatic Variant, aggiungere i file manifest utilizzando la funzione Module Settings (Impostazioni modulo). Questa funzione richiede i permessi a livello di amministratore. Per maggiori informazioni, vedere la *Guida di consultazione del NextSeq 550Dx Instrument (documento n. 1000000009513)*.

Specifica dei campioni per la corsa


Specificare i campioni per la corsa mediante una delle seguenti opzioni e procedure.

- ▶ **Enter samples manually** (Immissione manuale dei campioni): utilizzare la tabella vuota che si trova nella schermata Create Run (Crea corsa).
- ▶ **Import samples** (Importazione dei campioni): individuare un file esterno contenente valori separati da virgola (formato *.csv). Dalla schermata Create Run (Crea corsa) è possibile scaricare un modello.

Immissione manuale dei campioni

- 1 Immettere un nome del campione univoco (**modulo di analisi Somatic Variant**) o ID campione (**modulo di analisi Germline Variant**).
Utilizzare caratteri alfanumerici, trattini o trattini bassi.
- 2 **[Facoltativo]** Per i campioni di controllo positivi o negativi, fare clic con il pulsante destro del mouse e selezionare il tipo di controllo.
Il campione di controllo di un campione popola automaticamente il pozzetto corrispondente nell'altro raggruppamento in pool assegnando lo stesso campione di controllo.
- 3 **[Facoltativo]** Immettere una descrizione del campione nel campo Sample Description (Descrizione del campione).
Utilizzare caratteri alfanumerici, trattini o trattini bassi.
- 4 Selezionare un adattatore di indice 1 dall'elenco a discesa Index 1 (i7) (Indice 1 - i7).
Quando si utilizzano i pozzetti campione suggeriti, il software popola automaticamente gli adattatori di indice i7 e i5 che soddisfano i requisiti di diversità d'indice. Se il numero esatto di campioni da analizzare non compare nell'elenco, assicurarsi di selezionare gli adattatori di indice per i pozzetti in più.
- 5 Selezionare un adattatore di indice 2 dall'elenco a discesa Index 2 (i5) (Indice 2 - i5).
- 6 Selezionare un file manifest dall'elenco a discesa Manifest (File manifest).
I campioni in Pool A (Raggruppamento A) richiedono un file manifest diverso rispetto ai campioni in Pool B (Raggruppamento B).
- 7 Scegliere un'opzione per visualizzare, stampare o salvare il layout della piastra da utilizzare come riferimento al momento della preparazione delle librerie:
 - Selezionare l'icona  **Print** (Stampa) per visualizzare il layout della piastra. Selezionare **Print** (Stampa) per stampare il layout della piastra.
 - Selezionare **Export** (Esporta) per esportare le informazioni sui campioni in un file esterno.
- 8 Selezionare **Save Run** (Salva corsa).

Importazione dei campioni

- 1 Selezionare **Import Samples** (Importa campioni) e aprire il percorso del file contenente le informazioni relative al campione. È possibile importare due tipi di file:
 - Selezionare **Template** (Modello) sulla schermata Create Run (Crea corsa) per creare un nuovo layout della piastra. Il file modello contiene le intestazioni di colonna corrette per eseguire l'importazione. In ciascuna colonna, immettere le informazioni sui campioni da analizzare nella corsa. Eliminare le informazioni di esempio nelle caselle non utilizzate, quindi salvare il file.
 - Utilizzare un file contenente le informazioni sui campioni, che era stato esportato dal modulo Germline Variant o Somatic Variant mediante la funzione Export (Esporta).
- 2 Selezionare l'icona  **Print** (Stampa) per visualizzare il layout della piastra.
- 3 Selezionare **Print** (Stampa) per stampare il layout della piastra da utilizzare come riferimento per la preparazione delle librerie.

- 4 Selezionare **Save Run** (Salva corsa).

Preparazione della cartuccia di reagenti

Attenersi scrupolosamente alle seguenti istruzioni per la cartuccia di reagenti per ottenere un sequenziamento corretto.

- 1 Rimuovere la cartuccia di reagenti dal luogo di conservazione con temperatura tra -25 °C e -15 °C.
- 2 Scegliere uno dei seguenti metodi per scongelare i reagenti. Non sommergere la cartuccia. Una volta scongelata la cartuccia, asciugarla prima di passare alla fase successiva.

Temperatura	Tempo di scongelamento	Limite di stabilità
Bagno d'acqua tra 15 °C e 30 °C	60 minuti	Non superare le sei ore
Tra 2 °C e 8 °C	7 ore	Non superare i cinque giorni

NOTA Se si scongela più di una cartuccia nello stesso bagno d'acqua, consentire ulteriore tempo di scongelamento.

- 3 Capovolgere la cartuccia cinque volte per miscelare i reagenti.
- 4 Ispezionare la parte inferiore della cartuccia per assicurarsi che i reagenti siano scongelati e privi di precipitati. Confermare che le posizioni 29, 30, 31 e 32 siano scongelate, in quanto sono le più grandi e impiegano un tempo di scongelamento superiore.
- 5 Picchiettare delicatamente sul banco per ridurre le bolle d'aria.
Per risultati ottimali, procedere direttamente caricando il campione e impostando la corsa.

Preparazione della cella a flusso

- 1 Rimuovere dalla scatola una nuova cella a flusso dalla temperatura di conservazione compresa tra 2 °C e 8 °C.
- 2 Rimuovere la confezione in alluminio dalla scatola e tenerla a temperatura ambiente per 30 minuti.

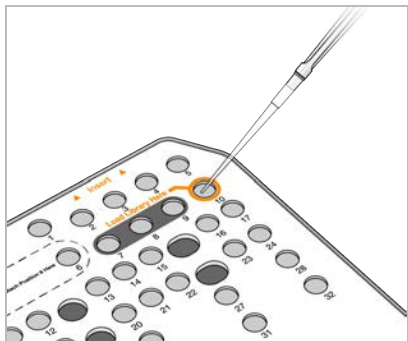
Preparazione delle librerie per il sequenziamento

Denaturare e diluire le librerie a un volume di caricamento di 1,3 ml. In pratica, la concentrazione di caricamento può variare in base ai metodi di preparazione e di quantificazione delle librerie. La diluizione delle librerie di campioni dipende dalla complessità dei raggruppamenti in pool di oligonucleotidi. Per le istruzioni sulla preparazione delle librerie dei campioni per il sequenziamento, inclusi la diluizione delle librerie e il raggruppamento in pool, vedere la sezione Istruzioni per l'uso applicabile al kit di preparazione delle librerie. È richiesta l'ottimizzazione della densità dei cluster su NextSeq 550Dx.

Caricamento delle librerie sulla cartuccia di reagenti

- 1 Pulire il sigillo in alluminio che copre il serbatoio n. 10 etichettato **Load Library Here** (Carica qui le librerie) con un panno a bassissimo rilascio di particelle.
- 2 Forare il sigillo con la punta di una pipetta pulita da 1 ml.
- 3 Caricare 1,3 ml di librerie preparate nel serbatoio n. 10 etichettato **Load Library Here** (Carica qui le librerie). Non toccare il sigillo in alluminio mentre si dispensano le librerie.

Figura 1 Caricamento delle librerie



Impostazione di una corsa di sequenziamento

- 1 Accedere a NextSeq 550Dx con la password del software Local Run Manager.
- 2 Dalla schermata Home (inizio) del software NOS, selezionare **Sequence** (Sequenziamento).
- 3 Selezionare una corsa dall'elenco, quindi selezionare **Next** (Avanti).
Si apre una serie di schermate per l'impostazione della corsa nell'ordine seguente: Load Flow Cell (Caricamento cella a flusso), Load Buffer Cartridge (Caricamento della cartuccia di tampone), Load Reagent Cartridge (Caricamento della cartuccia di reagenti) e Pre-run Check (Verifica pre-corsa).
- 4 Quando viene visualizzata la schermata Load Flow Cell (Caricamento cella a flusso), pulire e caricare la cella a flusso.
 - ▶ Rimuovere la cella a flusso dalla confezione in alluminio.
 - ▶ Aprire la confezione in plastica trasparente a forma di conchiglia e rimuovere la cella a flusso.
 - ▶ Pulire la superficie in vetro della cella a flusso con una salvietta imbevuta di alcool che non lascia residui. Asciugare il vetro con un panno da laboratorio a bassissimo rilascio di particelle.
 - ▶ Assicurarsi che la superficie in vetro della cella a flusso sia pulita. Se necessario, ripetere la fase di pulizia.
 - ▶ Rimuovere la cella a flusso usata in una corsa precedente.
 - ▶ Utilizzare i perni di allineamento per posizionare la cella a flusso sul piano portacelle.
- 5 Selezionare **Load** (Carica).
Lo sportello si chiude automaticamente, l'ID della cella a flusso viene visualizzato sulla schermata e i sensori sono sottoposti a verifica.
- 6 Attenersi alle indicazioni del software per svuotare il contenitore dei reagenti usati, caricare la cartuccia di tamponi NextSeq 550Dx e caricare la cartuccia di reagenti NextSeq 550Dx.
Una volta caricate le cartucce di tampone e di reagenti NextSeq 550Dx, il software legge e registra gli identificatori a radio frequenza (Radio-Frequency IDentification, RFID). Gli ID delle cartucce di tampone e di reagenti vengono visualizzati sullo schermo e i sensori sottoposti a verifica.
- 7 Al termine della verifica pre-corsa automatica, selezionare **Start** (Avvia). La selezione non è richiesta se la verifica è stata impostata per l'avvio automatico.
- 8 La schermata Sequencing (Sequenziamento) si apre all'inizio della corsa. La schermata fornisce una rappresentazione visiva dell'avanzamento della corsa, che comprende le intensità e i punteggi qualitativi (Q-scores).

Risultati

Real-Time Analysis (RTA) è un software integrato che esegue l'analisi delle immagini e l'identificazione delle basi e assegna un punteggio qualitativo a ciascuna base per ciascun ciclo di sequenziamento. Al termine dell'analisi

primaria, il modulo Local Run Manager selezionato sullo strumento NextSeq 550Dx avvia automaticamente l'analisi secondaria. Il processo di analisi secondaria qui descritto si riferisce ai moduli Germline Variant e Somatic Variant.

Demultiplex

Il demultiplex confronta ogni sequenza Index Read (Lettura indici) sulle sequenze d'indice specificate per la corsa. In questa fase non vengono considerati i valori qualitativi.

Le letture indici sono identificate mediante le fasi successive:

- ▶ I campioni sono numerati a partire da 1 in base all'ordine in cui sono stati elencati per la corsa.
- ▶ Il numero del campione 0 è riservato ai cluster non assegnati a un campione.
- ▶ I cluster sono assegnati a un campione quando la sequenza d'indice corrisponde esattamente o quando è presente una sola mancata corrispondenza per Index Read (Lettura indici).

Generazione di file FASTQ

Al termine del demultiplex, il software genera file dell'analisi intermedia in formato FASTQ, un formato di testo utilizzato per rappresentare le sequenze. I file FASTQ contengono le letture per ogni campione e i punteggi qualitativi associati. I cluster che non hanno superato il filtro sono esclusi.

Ogni file FASTQ contiene le letture per un solo campione e il nome di quel campione è incluso nel nome del file FASTQ. Nei moduli Germline Variant e Somatic Variant, vengono generati otto file FASTQ per ogni campione di ogni raggruppamento in pool di oligonucleotidi, quattro da Read 1 (Lettura 1) e quattro da Read 2 (Lettura 2). In totale, quindi si avranno 8 e 16 file FASTQ per ogni campione per Germline Variant e Somatic Variant, rispettivamente. I file FASTQ rappresentano gli input principali per l'allineamento.

Allineamento

Durante la fase di allineamento, l'algoritmo con matrice a banda di Smith-Waterman allinea i cluster ottenuti da ciascun campione sulle sequenze degli ampliconi specificati nel file manifest.

L'algoritmo Smith-Waterman con matrice a bande esegue gli allineamenti semi-globali delle sequenze per determinare regioni simili tra due sequenze. Piuttosto che confrontare la sequenza intera, l'algoritmo di Smith-Waterman confronta i segmenti di tutte le lunghezze possibili.

Ogni lettura paired-end viene valutata inizialmente in termini di allineamento con le sequenze sonda pertinenti per quella lettura.

- ▶ Read 1 (Lettura 1) è valutata rispetto al complemento inverso degli oligonucleotidi specifici per il locus a valle (Downstream Locus Specific Oligo, DLSO).
- ▶ Read 2 (Lettura 2) è valutata rispetto agli oligonucleotidi specifici per il locus a monte (Upstream Locus-Specific Oligo, ULSO).
- ▶ Se l'inizio di una lettura corrisponde a una sequenza sonda con non più di una non corrispondenza, tutta la lunghezza della lettura viene allineata rispetto all'amplicone target per quella sequenza.
- ▶ Se l'inizio di una lettura corrisponde a una sequenza sonda che non presenta più di tre differenze (mancate corrispondenze o variazioni causate da Indel anticipate), l'intera lunghezza della lettura viene allineata rispetto all'amplicone target per quella sequenza.
- ▶ Data la chimica del saggio, non si osservano Indel né nei DLSO né negli ULSO.

Gli allineamenti sono filtrati sui risultati degli allineamenti basati sulle percentuali di mancate corrispondenze sulla regione di interesse o sull'intero amplicone, in base alla lunghezza dell'amplicone. Gli allineamenti filtrati vengono scritti nei file di allineamento come non allineati e non vengono utilizzati per l'identificazione delle varianti.

Identificazione delle varianti

L'identificatore delle varianti Pisces è progettato per identificare le varianti SNV e le Indel da librerie preparate per lo strumento.

Report e ulteriori file di output

I moduli di analisi per le varianti creano report in formato PDF e di testo delimitato da tabulazione (*.txt) che visualizzano metriche come la profondità del sequenziamento e i conteggi delle varianti. I moduli creano inoltre file di output come i file VCF e i file nel formato di identificazione delle varianti del genoma (Genome Variant Call Format, gVCF).

Procedure per il controllo qualità

Il software NextSeq 550Dx valuta ciascuna corsa, campione e identificazione delle basi rispetto a metriche di controllo qualità. Nella preparazione delle librerie sono inoltre raccomandati controlli positivi e negativi e questi dovrebbero essere valutati. Valutare i controlli nel modo seguente:

- **Controllo negativo (controllo non templatato) o altri controlli negativi:** devono generare il risultato previsto. Se il controllo negativo genera un risultato diverso da quello previsto, significa che si è verificato un possibile errore nel monitoraggio del campione, una registrazione errata dei primer di indicizzazione o una contaminazione.
- **Campione di controllo positivo:** deve generare il risultato previsto. Se il controllo positivo genera un risultato diverso da quello previsto significa che si è verificato un possibile errore nel monitoraggio del campione oppure una registrazione errata dei primer di indicizzazione.

Caratteristiche delle prestazioni

Le caratteristiche delle prestazioni per lo strumento NextSeq 550Dx sono state stabilite utilizzando i moduli Germline Variant e Somatic Variant con TruSeq Custom Amplicon Kit Dx e NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e confermati utilizzando NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Gli studi hanno incluso indicizzazione dei campioni, carry-over dei campioni, input di DNA, sensibilità analitica (limite del campione bianco/limite del rilevamento), accuratezza, precisione, confronto del metodo e riproducibilità.

Gli studi analitici con NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) sono stati progettati per valutare le affermazioni sulle prestazioni precedentemente stabilite con NextSeq 550Dx High Output Reagent Kit v2 (300 cycles). I risultati dimostrano che, utilizzando TruSeq Custom Amplicon Kit Dx, i kit di reagenti (v2 e v2.5) hanno prestazioni comparabili. Per le caratteristiche prestazionali relative ai fattori pre-analitici, come i metodi di estrazione o le sostanze interferenti, vedere l'insero della confezione di *TruSeq Custom Amplicon Kit Dx*.

Definizione dei calcoli utilizzati nelle caratteristiche delle prestazioni

- 1 La concordanza positiva espressa in percentuale (Positive Percent Agreement, PPA) è calcolata come la proporzione dei loci classificati come varianti mediante un metodo di riferimento riportati correttamente dal saggio.
 - ▶ $(n. \text{ di loci delle varianti riportati correttamente dal saggio}) / (n. \text{ totale di loci delle varianti})$
I loci delle varianti riportati dal saggio che concordano con il metodo di riferimento sono veri positivi (True Positive, TP). I loci delle varianti riportati dal saggio come identificazioni di riferimento o come identificazioni delle varianti diverse sono falsi negativi (False Negative, FN).
- 2 La concordanza negativa espressa in percentuale (Negative Percent Agreement, NPA) è calcolata come la proporzione dei loci classificati come wild type mediante un metodo di riferimento riportati correttamente dal saggio.
 - ▶ $(n. \text{ di loci wild type riportati correttamente dal saggio}) / (n. \text{ totale di loci wild type})$
I loci wild type riportati dal saggio che concordano con il metodo di riferimento sono veri negativi (True Negative, TN). I loci wild type riportati dal saggio come varianti sono falsi positivi (False Positive, FP).

- 3 La concordanza complessiva espressa in percentuale (Overall Percent Agreement, OPA) è calcolata come la proporzione dei loci riportati correttamente dal saggio in base a un metodo di riferimento.
 - ▶ $[(n. \text{ di loci delle varianti riportato correttamente dal saggio}) + (n. \text{ di loci wild type riportato correttamente dal saggio})] / [(n. \text{ totale di loci delle varianti}) + (n. \text{ totale di loci wild type})]$
- 4 I calcoli di PPA, NPA e OPA non includono le identificazioni non rilevate (i loci delle varianti o dei riferimenti che non soddisfano uno o più filtri di qualità).
- 5 La percentuale di identificazione autosomica è calcolata come il numero totale di loci che attraversano i filtri diviso per il numero totale delle posizioni sequenziate per i cromosomi 1–22; i cromosomi X e Y sono esclusi. Questa metrica non prende in considerazione la concordanza delle identificazioni rilevate dal metodo di riferimento.

NextSeq 550Dx High Output Reagent Kit v2 (300 cycles)

Indicizzazione dei campioni

Gli index primer del campione, aggiunti durante la preparazione delle librerie, assegnano una sequenza univoca a ciascun campione di DNA. Queste sequenze univoche consentono il raggruppamento in pool di più campioni in una singola corsa di sequenziamento. L'indicizzazione dei campioni viene utilizzata sia per il flusso di lavoro Germline che per il flusso di lavoro Somatic. Lo scopo di questo studio era stabilire il numero di campioni minimo (8) e massimo (96) che lo strumento NextSeq 550Dx è in grado di sequenziare in una singola corsa di sequenziamento. Sono stati analizzati otto campioni univoci Platinum Genome con 12 diverse combinazioni di primer di indicizzazione per campione. I risultati dei campioni ottenuti da quattro corse di sequenziamento utilizzando il modulo Germline Variant sono stati confrontati con Platinum Genomes versione 2016-1.0.

Per il primo set di corse, sono state analizzate 96 librerie di campioni indicizzati univocamente con un saggio rappresentativo progettato per interrogare diversi geni che coprono 12.588 basi per filamento su tutti i 23 cromosomi umani al fine di verificare la capacità del saggio di identificare i genotipi in modo coerente per un dato campione su diverse combinazioni di primer di indicizzazione. Per il secondo set di corse, sono state analizzate otto librerie di campioni indicizzati univocamente in due corse di sequenziamento al fine di verificare il numero minimo di indici supportati.

Per le corse a 96 indici, la PPA per SNV andava da 98,7% a 100%, la PPA per le inserzioni e le delezioni era del 100% e la NPA era del 100% per ogni combinazione di 96 indici. Le corse a otto indici presentavano valori di PPA del 100% (SNV, inserzioni e delezioni) e valori di NPA del 100% per ogni combinazione di otto indici.

Carry-over dei campioni

Lo strumento NextSeq 550Dx consente, in una singola corsa di sequenziamento, di sequenziare più campioni oltre ai campioni di controllo. È stato condotto uno studio per valutare la portata del carry-over dei campioni in una corsa di sequenziamento (entro una corsa) e tra corse di sequenziamento (da corsa a corsa). Sono stati analizzati due campioni Platinum Genome, uno maschile e uno femminile, con un saggio rappresentativo progettato per interrogare diversi geni che coprono 12.588 basi (150 ampliconi) su 23 diversi cromosomi, inclusi entrambi i cromosomi sessuali. Le librerie sono state sequenziate sullo strumento NextSeq 550Dx utilizzando il modulo Germline Variant. È stato osservato il carry-over dei campioni maschili nei campioni femminili in presenza di letture dell'amplicone del cromosoma Y nei campioni femminili.

Il carry-over entro la corsa potrebbe essere stato introdotto durante la generazione di cluster, l'identificazione delle basi del ciclo degli indici e il demultiplex campioni. Per analizzare il carry-over dei campioni entro una corsa di sequenziamento, sono stati sequenziati una volta un pool della libreria che consisteva di 46 replicati di ciascun campione maschile e femminile oltre a quattro controlli non templati sullo strumento NextSeq 550Dx. Il carry-over dei campioni entro una corsa è stato valutato confrontando la copertura dell'amplicone del cromosoma Y di ciascun replicato femminile sulla copertura media dell'amplicone del cromosoma Y di tutti i replicati maschili presenti nel raggruppamento in pool. La mediana osservata del carry-over entro la corsa era di 0,084%.

Per analizzare il carry-over dei campioni da corsa a corsa, sono stati preparati due raggruppamenti in pool delle librerie e sequenziati su un solo strumento NextSeq 550Dx. Il primo raggruppamento in pool conteneva 46 replicati del campione femminile oltre a due campioni di controllo non templato. Il secondo raggruppamento

in pool conteneva 46 replicati del campione maschile oltre a due campioni di controllo non templatato. Entrambi i raggruppamenti in pool hanno utilizzato il medesimo set di adattatori indici. Il raggruppamento in pool del campione femminile è stato sequenziato per primo, di seguito è stata eseguita una corsa di sequenziamento con il raggruppamento in pool del campione maschile quindi un'altra corsa di sequenziamento ripetuta del raggruppamento in pool del campione femminile. Il carry-over dei campioni da corsa a corsa è stato valutato confrontando la copertura dell'amplicone del cromosoma Y tra i replicati corrispondenti ottenuti dalla corsa ripetuta del raggruppamento in pool del campione femminile e dalla corsa del raggruppamento in pool del campione maschile. La mediana osservata del carry-over da corsa a corsa era di 0,0076%.

Input di DNA

Sangue (Germline)

Per lo strumento NextSeq 550Dx è stato stabilito l'intervallo di input di DNA di sangue per la preparazione delle librerie con TruSeq Custom Amplicon Kit Dx utilizzando il flusso di lavoro per il modulo Germline Variant. L'intervallo è stato valutato eseguendo uno studio di diluizione in serie utilizzando 13 campioni Platinum Genome con un saggio rappresentativo progettato per investigare diversi geni che coprono 12.588 basi su 23 diversi cromosomi. La libreria è stata sequenziata su due strumenti NextSeq 550Dx utilizzando un lotto di NextSeq 550Dx High Output Reagent Kit v2 (300 cycles).

Cinque campioni sono stati analizzati in duplicati a cinque livelli di input di DNA che andavano da 250 ng a 12 ng (250 ng, 100 ng, 50 ng, 25 ng e 12 ng). Otto campioni sono stati analizzati come un singolo replicato a ciascuno dei cinque livelli di input di DNA. Per determinare l'accuratezza, i genotipi dei campioni sono stati confrontati con Platinum Genomes versione 2016-1.0. I risultati sono stati determinati per ciascun livello di input. La PPA per ciascun tipo di variante (SNV, inserzioni e delezioni) è rappresentata nella [Tabella 1](#); la NPA è rappresentata nella [Tabella 2](#). Tutti i livelli di input presentano un'accuratezza simile. L'input di DNA raccomandato per TruSeq Custom Amplicon Kit Dx è di 50 ng; 25 ng e 100 ng forniscono il limite inferiore e il limite superiore per soddisfare le caratteristiche delle prestazioni.

Tabella 1 Risultati della PPA per ciascun input di DNA per tipo di variante

Input di DNA (ng)	Tipo di variante	Varianti previste	TP	FN	Identificazioni varianti non rilevate	PPA (%)
12	SNV	2.412	2.381	31	0	98,7
25			2.404	8	0	99,7
50			2.403	9	0	99,6
100			2.412	0	0	100
250			2.412	0	0	100
12	Inserzione	808	784	3	21	99,6
25			781	5	22	99,4
50			786	2	20	99,8
100			786	0	22	100
250			786	0	22	100
12	Delezione	758	732	12	14	98,4
25			737	7	14	99,1
50			742	2	14	99,7
100			744	0	14	100
250			744	0	14	100

Tabella 2 NPA per ciascun input di DNA

Input di DNA (ng)	TN	FP	Identificazioni riferimenti non rilevate	NPA (%)
12	430.940	4	26	> 99,9
25	430.936	0	34	100
50	430.936	2	32	> 99,9
100	430.942	0	28	100
250	430.942	0	28	100

FFPE (Somatic)

Per lo strumento NextSeq 550Dx è stato stabilito l'intervallo di input di DNA fissato in formalina e incluso in paraffina (Formalin-Fixed Paraffin-Embedded, FFPE) per la preparazione delle librerie con TruSeq Custom Amplicon Kit Dx utilizzando il flusso di lavoro per il modulo Somatic Variant. L'intervallo di input di DNA è stato valutato eseguendo uno studio di diluizione in serie utilizzando tre campioni Platinum Genome con un saggio rappresentativo progettato per investigare diversi geni che coprono 12.588 basi su 23 diversi cromosomi. Prima dell'estrazione del DNA, le linee cellulari GM12878 e GM12877 Platinum Genome sono state fissate in formalina e incluse in paraffina. GM12878 è stato diluito con GM12877 in modo da ottenere frequenze alleliche delle varianti (Variant Allele Frequency, VAF) di 81 varianti (55 SNV, 10 inserzioni e 16 delezioni) prossime a 0,025, 0,05 o 0,10. Inoltre ogni campione presentava 91 varianti con frequenze di varianti superiori, fino a 1,0 VAF. I campioni sono stati elaborati in duplicati ai cinque livelli di input di DNA con il ciclo quantitativo delta medio (Delta Quantitative Cycle, dCq) di 2,1, 3,6, 4,6, 6,0 e 7,8 come misurato da TruSeq Custom Amplicon Dx - FFPE QC Kit. Ciascuna libreria è stata sequenziata su due strumenti NextSeq 550Dx utilizzando due lotti di NextSeq 550Dx High Output Reagent Kit v2 (300 cycles). Per determinare l'accuratezza, le identificazioni delle varianti dei campioni sono state confrontate con Platinum Genomes versione 2016-1.0. La PPA per ciascun tipo di variante (SNV, inserzioni e delezioni) è rappresentata nella [Tabella 3](#); la NPA è rappresentata nella [Tabella 4](#). L'input di DNA raccomandato per le varianti a una VAF pari o superiore a 0,05 è dCq \leq 4. 4,6 fornisce il limite inferiore per soddisfare le caratteristiche delle prestazioni.

Tabella 3 Risultati della PPA per ciascun input di DNA per tipo di variante

dCq media	Tipo di variante	Varianti previste	Identificazioni non rilevate previste	VAF diluizione target					
				0,025		0,05		0,10	
				Identificazioni varianti non rilevate	PPA (%)	Identificazioni varianti non rilevate	PPA (%)	Identificazioni varianti non rilevate	PPA (%)
2,1	SNV	808	Non applicabile.	196	100	0	100	0	100
3,6				250	99,3	4	100	0	100
4,6				251	94,6	51	99,2	5	100
6,0				257	65,3	213	91,4	100	100
7,8				254	69,3	185	90,7	100	100
2,1	Inserzione	264	8	66	96,5	8	100	8	100
3,6				62	97,0	8	100	8	100
4,6				48	96,3	21	100	8	100
6,0				40	80,4	47	98,2	24	95,8
7,8				57	87,0	56	96,2	31	100

2,1	Delezione	304	16	58	100	16	100	16	100
3,6				80	100	16	100	16	100
4,6				65	95,4	28	100	16	100
6,0				78	74,8	105	94,0	36	100
7,8				76	75,0	79	95,1	57	98,8

Tabella 4 NPA per ciascun input di DNA

dCq media	Wild type previsto	VAF diluizione target					
		0,025		0,05		0,10	
		Identificazioni riferimenti non rilevate	NPA (%)	Identificazioni riferimenti non rilevate	NPA (%)	Identificazioni riferimenti non rilevate	NPA (%)
2,1	93.688	344	100	260	100	324	100
3,6		400	100	332	100	380	100
4,6		1.308	100	1.336	100	784	100
6,0		3.900	> 99,9	3.296	> 99,9	2.996	100
7,8		3020	> 99,9	2.880	> 99,9	2.448	> 99,9

Sensibilità analitica: limite del campione bianco (Limit of Blank, LoB) e limite del rilevamento (Limit of Detection, LoD)

Questo studio è stato condotto per valutare il limite del campione bianco (LoB) e il limite del rilevamento (LoD) per il modulo Somatic Variant sullo strumento NextSeq 550Dx. Lo studio è stato eseguito utilizzando un saggio rappresentativo progettato per interrogare diversi geni che coprono 12.588 basi su 23 diversi cromosomi. Prima dell'estrazione del DNA, le linee cellulari GM12878 e GM12877 Platinum Genome sono state fissate in formalina e incluse in paraffina. GM12878 è stato diluito con GM12877 in modo da ottenere frequenze delle varianti per 74 varianti (53 SNV, 7 inserzioni e 14 delezioni) pari a $0,05 \pm 0,02$. GM12877 e GM12878 (GM12878-D) diluito sono stati analizzati per sei giorni di avvio consecutivi con un singolo strumento, utilizzando alternativamente due lotti di NextSeq 550Dx High Output Reagent Kit v2 (300 cycles), per un totale di sei corse di sequenziamento. In questo modo sono stati ottenuti 60 replicati per ciascuna variante in GM12878-D e 72 replicati per ciascuna coordinata wild type corrispondente in GM12877 per ciascun lotto di reagenti. LoB e LoD sono stati calcolati con l'approccio classico indicato da CLSI EP17-A2 utilizzando l'opzione non parametrica. LoB e LoD sono stati calcolati per SNV, inserzioni e delezioni separatamente raggruppando in pool le frequenze delle varianti per un dato tipo di variante. L'errore di tipo I era definito come 0,01 e l'errore di tipo II era definito come 0,05.

Per LoB, le frequenze delle varianti raggruppate in pool sono state ordinate dalla più bassa alla più alta ed è stata calcolata la posizione 99 per ciascun lotto di reagenti per ciascun tipo di variante (Tabella 5). Il modulo Somatic Variant utilizza un valore di cutoff (l'effettivo LoB) di 0,026 di VAF per determinare il rilevamento qualitativo delle varianti. I limiti di LoB calcolati hanno verificato che questo cutoff ha comportato un errore di tipo I di non più di 0,01.

Tabella 5 Limite del bianco

Tipo di variante	Osservazioni totali	LoB del lotto di reagenti 1 (%)	LoB del lotto di reagenti 2 (%)
SNV	3.816	0,77	0,77
Inserzione	504	0,56	0,56
Delezione	1.008	1,20	1,20

Per LoD, è stata calcolata la percentuale della frequenza delle singole mutazioni per ciascun lotto di reagenti per ciascun tipo di variante che risultava al di sotto del valore di cutoff di 0,026 Tabella 6. Poiché le percentuali erano inferiori rispetto all'errore di tipo II del 5% (0,05), la mediana delle frequenze combinate delle varianti sono state

calcolate come il valore LoD (Tabella 6). Il valore LoD per ciascun tipo di variante è stato preso come il più grande dei due valori calcolati per i due lotti di reagenti: 4,97% per le SNV, 5,12% per le inserzioni e 5,26% per le delezioni.

Tabella 6 Limite del rilevamento

Lotto di reagenti	Tipo di variante	Osservazioni totali	N. di misurazioni VAF < 2,6%	% di misurazioni VAF < 2,6%	Limite del rilevamento (%)
1	SNV	3.180	53	1,7	4,94
	Inserzione	420	6	1,4	5,08
	Delezione	840	7	0,8	5,22
2	SNV	3.180	51	1,6	4,97
	Inserzione	420	5	1,2	5,12
	Delezione	840	7	0,80	5,26

Accuratezza

Germline

Il seguente studio è stato condotto per valutare l'accuratezza dell'identificazione delle varianti del modulo Germline Variant sullo strumento NextSeq 550Dx utilizzando NextSeq 550Dx High Output Reagent Kit v2 (300 cycles). Sono stati analizzati 13 campioni univoci Platinum Genome utilizzando un saggio rappresentativo progettato per interrogare diversi geni che coprono 12.588 basi (150 ampliconi) su 23 diversi cromosomi. È stato eseguito un totale di nove corse utilizzando tre strumenti di sequenziamento, tre lotti di reagenti e tre operatori su cinque giorni di avvio. L'accuratezza è stata determinata per SNV, inserzioni e delezioni confrontando i risultati sul metodo di riferimento composito ben caratterizzato, Platinum Genomes versione 2016-1.0. Le regioni genomiche affidabili sono state definite in base a questo metodo di riferimento, se non altrimenti specificato.

Tabella 7 Riepilogo della concordanza per Germline

Criteri	Osservazioni totali ¹	Risultato per osservazione ²	Risultato per corsa ³
PPA per SNV	819	98,7	> 99,9
PPA per le inserzioni	819	95,0	98,9
PPA per le delezioni	819	100	100
NPA	819	100	100
OPA	819	> 99,9	> 99,9

¹ Calcolato come il numero di campioni per corsa (91) x il numero di corse (9) = 819.

² Il valore più basso osservato per replicato di campione su tutte le nove corse.

³ Il valore più basso quando i dati di ciascuna corsa sono stati analizzati in aggregato.

La Tabella 8 contiene i dati dello studio con la concordanza positiva e negativa espressa in percentuale in base ai singoli campioni, dove, per i calcoli della PPA, i risultati delle varianti sono confrontati con Platinum Genomes versione 2016-1.0. Sono stati combinati i tre tipi di varianti (SNV, inserzioni e delezioni). Poiché il metodo di riferimento fornisce solo risultati per le varianti di singolo nucleotide e inserzioni/delezioni, i risultati delle basi non varianti sono confrontati con la sequenza di riferimento del genoma umano versione hg19, per i calcoli della NPA.

Tabella 8 Concordanza per campione per Germline

Campione	Percentuale di identificazione media	Varianti previste ¹	TP	FN	Identificazioni varianti non rilevate	TN	FP	PPA	NPA	OPA
NA12877	> 99,9	4.788	4.788	0	0	756.762	0	100	100	100
NA12878	> 99,9	8.505	8.379	1	125	751.464	0	> 99,9	100	> 99,9
NA12879	> 99,9	6.048	5.985	5	58	757.701	0	99,9	100	> 99,9
NA12880	> 99,9	6.993	6.930	0	63	757.638	0	100	100	100
NA12881	> 99,9	7.875	7.811	3	61	751.653	0	> 99,9	100	> 99,9
NA12882	> 99,9	6.300	6.174	3	123	754.803	0	> 99,9	100	> 99,9
NA12883	> 99,9	7.119	7.056	0	63	751.905	0	100	100	100
NA12884	> 99,9	7.182	7.119	6	57	754.146	0	99,9	100	> 99,9
NA12885	> 99,9	7.686	7.560	2	124	754.173	0	> 99,9	100	> 99,9
NA12886	> 99,9	7.245	7.182	7	56	752.469	0	99,9	100	> 99,9
NA12887	> 99,9	7.119	7.119	0	0	750.645	0	100	100	100
NA12888	> 99,9	6.804	6.804	0	0	756.065	0	100	100	100
NA12893	> 99,9	7.434	7.371	1	62	750.015	0	> 99,9	100	> 99,9

¹ Il numero totale delle varianti in tutti i replicati dei campioni sulle nove corse.

La **Tabella 9** contiene i dati dello studio in base ai singoli campioni, dove i risultati delle varianti sono confrontati con un metodo di riferimento composito ben caratterizzato. Il rilevamento è stato valutato per ciascun tipo di variante: SNV, inserzioni e delezioni, separatamente. Sono state escluse le posizioni dei riferimenti.

Tabella 9 Concordanza per campione in base al tipo di variante per Germline

> di campione	SNV			Inserzioni			Delezioni		
	> di previsto	> di TP	> di FN	> di previsto	> di TP	> di FN	Previsto	TP	FN
NA12877	2.331	2.331	0	1.323	1.323	0	1.134	1.134	0
NA12878	5.733	5.733	0	1.260	1.197	1	1.512	1.449	0
NA12879	3.591	3.591	0	1.323	1.260	5	1.134	1.134	0
NA12880	4221	4221	0	1.512	1.512	0	1.260	1.197	0
NA12881	4914	4.913	1	1.512	1.449	2	1.449	1.449	0
NA12882	3717	3717	0	1.386	1.323	3	1.197	1.134	0
NA12883	4.284	4.284	0	1.449	1.449	0	1.386	1.323	0
NA12884	4.284	4.284	0	1.575	1.512	6	1.323	1.323	0
NA12885	4.725	4.725	0	1.575	1.512	2	1.386	1.323	0
NA12886	4.347	4.347	0	1.449	1.386	7	1.449	1.449	0
NA12887	4.284	4.284	0	1.323	1.323	0	1.512	1.512	0
NA12888	4.158	4.158	0	1.449	1.449	0	1.197	1.197	0
NA12893	4.599	4.599	0	1.386	1.323	1	1.449	1.449	0

I campioni sono stati ulteriormente analizzati per identificare piccole inserzioni e delezioni (Indel). Il riepilogo complessivo viene presentato nella **Tabella 10**. Era presente un totale di 71 Indel con una dimensione compresa nell'intervallo 1–24 bp per le inserzioni e 1–25 bp per le delezioni.

Tabella 10 Riepilogo del rilevamento delle Indel per Germline

Tipo di variante	Varianti previste	TP	FN	Identificazioni varianti non rilevate	PPA
Inserzione	18.522	18.018	27	477	99,9
Delezione	17.388	17.073	0	315	100

Il saggio rappresentativo consisteva di 150 ampliconi progettati per coprire diverso contenuto genomico. Il contenuto in GC degli ampliconi rientrava nell'intervallo compreso tra 0,19 e 0,87. Gli ampliconi presentavano inoltre un intervallo di ripetizioni di singolo nucleotide (ad es., PolyA, PolyT), dinucleotide e trinucleotide. I dati sono stati compilati in base ai singoli ampliconi (Tabella 11) per determinare gli effetti del contenuto genomico sulla percentuale di identificazioni corrette. La percentuale di identificazioni corrette consiste di identificazioni delle varianti e dei riferimenti ed è inferiore al 100% se sono presenti identificazioni errate o identificazioni non rilevate.

Tabella 11 Accuratezza a livello di amplicone per Germline

Amplicone	Cromosoma	Avvio amplicone	Fine amplicone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplicone	Contenuto in GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette
1	1	36450499	36450591	93	93	Indel	0,22	76.167	0	0	100
2	1	109465122	109465200	79	79	Poly A (5), Poly C (5), Indel	0,38	64.701	0	0	100
3	1	218353867	218353957	91	91	Indel	0,4	74.529	0	0	100
4	1	223906657	223906748	92	92	Indel	0,49	75.348	0	0	100
5	1	228526602	228526682	81	81	Poly G (5)	0,69	66.339	0	0	100
6	1	236372039	236372108	70	70	Poly T (10), Indel	0,39	57.330	0	0	100
7	1	247812041	247812128	88	88	Poly A (5), CT(3), TAA(3), Indel	0,27	72.072	0	0	100
8	2	55862774	55862863	90	90	Indel	0,28	73.710	0	0	100
9	2	87003930	87004009	80	80	Indel	0,38	65.520	0	0	100
10	2	177016721	177016805	85	81	N/A	0,65	66.339	0	0	100
11	2	186625727	186625801	75	75	Poly A (8)	0,35	61.425	0	0	100
12	2	190323504	190323591	88	88	Poly T (5)	0,42	72.072	0	0	100
13	2	200796740	200796826	87	87	Poly T (5), indel	0,31	71.253	0	0	100
14	2	212245049	212245139	91	91	Poly T (5), Poly A (6), Indel	0,3	74.529	0	0	100
15	2	228147052	228147144	93	93	Indel	0,43	76.167	0	0	100
16	2	235016350	235016422	73	73	Poly T (5), indel	0,42	59.787	0	0	100
17	3	4466229	4466321	93	93	AT(3), Indel	0,27	74.823	0	1.344	98,2
18	3	46620561	46620643	83	83	N/A	0,43	67.977	0	0	100
19	3	49851331	49851400	70	70	CT(3), Indel	0,49	57.330	0	0	100

Amplificone	Cromosoma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette
20	3	189713161	189713248	88	88	Poly A (5), Poly T (5), Poly A (9), TG(3)	0,41	72.072	0	0	100
21	3	190106030	190106104	75	74	Indel	0,57	60.543	0	63	99,9
22	4	2233667	2233744	78	78	Poly A (6)	0,26	63.882	0	0	100
23	4	7780541	7780637	97	97	Poly G (6), Poly T (5), Poly A (5)	0,42	79.443	0	0	100
24	4	15688604	15688681	78	78	N/A	0,29	63.882	0	0	100
25	4	56236521	56236586	66	62	Poly A (5), Indel	0,36	50.778	0	0	100
26	4	102839244	102839314	71	69	Poly A (5)	0,46	56.511	0	0	100
27	4	164446743	164446804	62	62	Poly A (7), Indel	0,27	50.778	0	0	100
28	5	1882081	1882158	78	75	N/A	0,78	61.425	0	0	100
29	5	14769061	14769144	84	84	GT(3), CCA(3)	0,62	68.796	0	0	100
30	5	41069808	41069871	64	64	N/A	0,39	52.416	0	0	100
31	5	74077114	74077196	83	83	Poly A (6), Indel	0,3	67.977	0	0	100
32	5	147475343	147475409	67	67	Poly T (5)	0,37	54.873	0	0	100
33	5	149323731	149323821	91	91	CT(4), AG(3)	0,55	74.529	0	0	100
34	5	155662213	155662287	75	75	Indel	0,43	61.425	0	0	100
35	6	6318713	6318814	102	102	Poly G (6)	0,68	83.538	0	0	100
36	6	24949983	24950074	92	92	Indel	0,63	75.348	0	0	100
37	6	31084900	31084999	100	94	GCT(5), Indel	0,61	76.608	0	378	99,5
38	6	32147987	32148084	98	98	Poly T (5), TCT(3), CTT(3)	0,55	80.262	0	0	100
39	6	32986864	32986958	95	95	Indel	0,53	77.805	0	0	100
40	6	33408498	33408583	86	86	Poly C (6)	0,7	70.434	0	0	100
41	6	41647401	41647495	95	94	Poly G (5), Indel	0,61	76.986	0	0	100
42	6	112435865	112435955	91	91	Poly A (5)	0,44	74.529	0	0	100
43	7	22202076	22202148	73	73	N/A	0,44	59.787	0	0	100

Amplificone	Cromosoma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette
44	7	66276100	66276187	88	88	Indel	0,35	72.072	0	0	100
45	7	77365735	77365821	87	87	Poly A (7), AG(4)	0,26	71.253	0	0	100
46	7	110939946	110940030	85	85	Indel	0,38	69.615	0	0	100
47	7	128533468	128533557	90	90	Poly G (5), Indel	0,62	73.710	0	0	100
48	7	149503875	149503965	91	91	Poly G (6), Poly C (6), Indel	0,71	74.529	0	0	100
49	7	154404519	154404599	81	66	N/A	0,31	54.054	0	0	100
50	7	156476507	156476599	93	93	Indel	0,35	76.167	0	0	100
51	8	1817312	1817394	83	83	N/A	0,42	67.977	0	0	100
52	8	24811020	24811109	90	89	Poly G (7), CTC(4), Indel	0,61	72.171	0	720	99,0
53	8	76518625	76518691	67	67	Indel	0,3	54.873	0	0	100
54	9	103054909	103055006	98	98	Poly G (6)	0,67	80.262	0	0	100
55	9	105586150	105586214	65	65	Indel	0,32	53.235	0	0	100
56	9	107620823	107620918	96	96	N/A	0,49	78.624	0	0	100
57	9	123769149	123769231	83	83	AT(3)	0,37	67.977	0	0	100
58	9	138995345	138995441	97	97	Poly C (6), Indel	0,68	79.443	0	0	100
59	10	5987120	5987198	79	78	Poly G (5), Indel	0,47	63.882	0	0	100
60	10	11784629	11784726	98	91	GC(3)	0,87	74.529	0	0	100
61	10	27317777	27317855	79	79	Poly T (5)	0,3	64.701	0	0	100
62	10	33018351	33018440	90	90	Poly A (5), Poly T (5)	0,2	73.710	0	0	100
63	10	45084159	45084253	95	95	Indel	0,35	77.805	0	0	100
64	10	55892599	55892687	89	88	AC(11), Indel	0,42	71.747	0	325	99,5
65	10	101611250	101611329	80	80	N/A	0,49	65.520	0	0	100
66	10	118351373	118351453	81	81	N/A	0,51	66.339	0	0	100
67	11	8159816	8159912	97	96	N/A	0,45	78.624	0	0	100

Amplificone	Cromosoma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette
68	11	30177648	30177717	70	70	Indel	0,46	57.330	0	0	100
69	11	47470345	47470444	100	100	N/A	0,65	81.900	0	0	100
70	11	59837679	59837740	62	62	Indel	0,37	50.778	0	0	100
71	11	64418856	64418957	102	102	N/A	0,59	83.538	0	0	100
72	11	93529612	93529684	73	73	Poly A (5)	0,4	59.787	0	0	100
73	11	101347052	101347136	85	85	N/A	0,42	69.615	0	0	100
74	11	102477336	102477426	91	91	Poly G (6)	0,55	74.529	0	0	100
75	11	118406285	118406369	85	85	Indel	0,53	69.615	0	0	100
76	11	120357801	120357885	85	85	Poly A (5), CA(3), Indel	0,34	69.615	0	0	100
77	11	125769313	125769397	85	85	GA(3)	0,52	69.615	0	0	100
78	12	2834770	2834853	84	84	Poly C (5), Indel	0,52	68.796	0	0	100
79	12	26811004	26811096	93	93	Poly A (7), AC(4)	0,33	76.167	0	0	100
80	12	30881766	30881846	81	81	N/A	0,49	66.339	0	0	100
81	12	88474105	88474175	71	71	Poly A (6)	0,35	58.149	0	0	100
82	12	120966872	120966966	95	95	Poly G (5)	0,68	77.805	0	0	100
83	13	24167504	24167576	73	73	N/A	0,52	59.787	0	0	100
84	13	25816961	25817049	89	88	Poly A (5), Poly T (7), Poly A (7), Indel	0,22	72.072	0	0	100
85	13	44880112	44880200	89	89	Indel	0,49	72.891	0	0	100
86	13	77665218	77665294	77	77	Indel	0,39	63.063	0	0	100
87	14	31619327	31619393	67	67	GA(3),TA(3)	0,39	54.873	0	0	100
88	14	39517884	39517966	83	83	N/A	0,25	67.977	0	0	100
89	14	46958962	46959034	73	72	Poly T (5), indel	0,19	58.642	0	326	99,4
90	14	58050030	58050110	81	81	Indel	0,38	66.339	0	0	100
91	14	82390559	82390649	91	91	Indel	0,35	74.529	0	0	100
92	14	92549544	92549609	66	66	Poly A (5)	0,41	54.054	0	0	100

Amplificone	Cromosoma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette
93	14	102808496	102808589	94	94	Indel	0,62	76.986	0	0	100
94	15	43170751	43170848	98	96	Poly C (5)	0,45	78.624	0	0	100
95	15	63446149	63446216	68	68	Indel	0,25	55.692	0	0	100
96	15	77879807	77879901	95	93	Poly G (5), Indel	0,68	76.167	0	0	100
97	15	81625334	81625428	95	95	Poly T (6)	0,43	77.805	0	0	100
98	15	85438263	85438334	72	71	Indel	0,65	58.149	0	0	100
99	15	89817413	89817503	91	91	N/A	0,36	74.529	0	0	100
100	15	89864274	89864343	70	70	Indel	0,56	57.330	0	0	100
101	16	1894910	1894972	63	63	N/A	0,27	51.597	0	0	100
102	16	28997904	28997998	95	95	Poly C (5)	0,67	77.805	0	0	100
103	16	53682908	53682994	87	87	TA(3)	0,41	71.253	0	0	100
104	16	57954406	57954509	104	104	Poly C (5)	0,67	85.176	0	0	100
105	16	85706375	85706465	91	91	Poly T (5), indel	0,37	74.529	0	0	100
106	17	3563920	3564008	89	89	GC(3)	0,64	72.891	0	0	100
107	17	3594191	3594277	87	87	Poly C (5), Indel	0,67	71.247	0	6	100
108	17	3970090	3970180	91	91	Indel	0,46	74.529	0	0	100
109	17	16084945	16085037	93	93	Indel	0,26	76.167	0	0	100
110	17	33998759	33998849	91	89	Poly T (5)	0,54	72.891	0	0	100
111	17	39589691	39589774	84	82	Poly A (13), Indel (x2)	0,29	66.343	27	788	98,8
112	17	41244394	41244484	91	91	Poly A (5)	0,34	74.529	0	0	100
113	17	45438866	45438957	92	92	Poly A (7), AT(3), AT(4), AT(4), Indel	0,26	75.348	0	0	100
114	17	61502432	61502510	79	79	Indel	0,41	64.413	0	288	99,6
115	17	64023582	64023667	86	86	Poly T (7)	0,22	70.434	0	0	100
116	17	72308237	72308320	84	84	GAG(3)	0,62	68.796	0	0	100
117	18	2616456	2616522	67	67	GA(3)	0,31	54.873	0	0	100

Amplificone	Cromosoma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette
118	18	6980478	6980568	91	91	N/A	0,37	74.529	0	0	100
119	18	9888026	9888094	69	69	Poly A (6), TG(3)	0,43	56.511	0	0	100
120	18	38836999	38837073	75	75	Poly A (5), Indel	0,37	61.425	0	0	100
121	18	47405382	47405462	81	81	CTC(3), Indel	0,47	66.339	0	0	100
122	18	54815665	54815749	85	85	CT(3), Indel	0,45	69.615	0	0	100
123	18	59773996	59774060	65	65	N/A	0,48	53.235	0	0	100
124	19	625143	625241	99	99	N/A	0,59	81.081	0	0	100
125	19	18121418	18121491	74	74	N/A	0,68	60.605	1	0	100
126	19	18186574	18186643	70	70	N/A	0,64	57.330	0	0	100
127	20	746056	746149	94	94	N/A	0,61	76.986	0	0	100
128	20	10633195	10633276	82	82	AC(3)	0,59	67.158	0	0	100
129	20	17705633	17705708	76	76	CT(3)	0,58	62.244	0	0	100
130	20	21766821	21766890	70	70	GT(3),TG(4), Indel	0,46	57.330	0	0	100
131	20	25278421	25278521	101	101	Indel	0,63	82.719	0	0	100
132	20	50897302	50897368	67	67	Indel	0,36	54.873	0	0	100
133	20	62331904	62331994	91	88	Poly G (6)	0,73	72.072	0	0	100
134	20	62690860	62690946	87	87	Indel	0,57	71.253	0	0	100
135	21	30300823	30300888	66	66	Indel	0,35	54.054	0	0	100
136	21	33694176	33694273	98	98	Poly T (6), CA(3)	0,54	80.262	0	0	100
137	21	36710706	36710792	87	87	GT(3), Indel	0,39	71.253	0	0	100
138	21	46644924	46644992	69	69	Poly A (6), AG(3), Indel	0,32	56.439	0	72	99,9
139	21	46705575	46705664	90	90	Poly T (5), Poly A (6)	0,5	73.710	0	0	100
140	22	25750774	25750873	100	100	Indel	0,63	81.900	0	0	100
141	22	32439233	32439329	97	97	N/A	0,68	79.443	0	0	100
142	22	37409844	37409940	97	97	Indel	0,46	79.443	0	0	100

Amplificone	Cromo- soma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identifi- cazioni corrette	Identifi- cazioni errate	Nessuna identifi- cazione	% di identifi- cazioni corrette
143	22	37637596	37637694	99	99	N/A	0,6	81.081	0	0	100
144	22	47081347	47081438	92	92	Indel	0,66	75.348	0	0	100
145	X	15870424	15870492	69	69	Poly T (5)	0,26	56.511	0	0	100
146	X	135288543	135288611	69	69	Poly C (5)	0,62	56.511	0	0	100
147	X	135290777	135290847	71	71	N/A	0,52	58.149	0	0	100
148	Si	2655397	2655461	65	0	N/A	0,55	0	0	0	N/A
149	Si	2655519	2655609	91	0	N/A	0,48	0	0	0	N/A
150	Si	2655609	2655679	71	0	Poly A (5)	0,37	0	0	0	N/A

I risultati del sequenziamento per il campione NA12878 sono stati confrontati con un genotipo altamente affidabile per NA12878, come stabilito dal National Institutes of Standards and Technology (NIST) (v.2.19). Dei 150 ampliconi, 92 ampliconi sono rientrati completamente nelle regioni genomiche altamente affidabili, 41 ampliconi presentavano una sovrapposizione parziale e 17 ampliconi non presentavano alcuna sovrapposizione nella sequenza NIST. Questo ha fornito 10.000 coordinate per replicato per il confronto. Le identificazioni delle basi non varianti sono state confrontate con la sequenza di riferimento del genoma umano versione hg 19. I risultati dell'accuratezza sono mostrati nella **Tabella 12**.

Tabella 12 Concordeza del campione NA12878 rispetto al database NIST per Germline

Campione	N. di ampliconi	Percentuale di identificazione media	TP	FN	TN	FP	PPA	NPA	OPA
NA12878	133	> 99,9	6.552	1	610.470	0	> 99,9	100	> 99,9

In base ai dati forniti da questo studio Germline di nove corse, lo strumento NextSeq 550Dx può sequenziare in modo coerente:

- ▶ Contenuto in GC $\geq 19\%$ (tutte le basi identificate in 819 ampliconi sequenziati con 19% di contenuto in GC identificato correttamente con una percentuale di identificazioni non rilevate di 0,6%)
- ▶ Contenuto in GC $\leq 87\%$ (tutte le basi identificate in 819 ampliconi sequenziati con 87% di contenuto in GC identificato correttamente con nessuna identificazione non rilevata)
- ▶ Lunghezze PolyA ≤ 9 (tutte le basi identificate in 819 ampliconi sequenziati contenenti una ripetizione PolyA di nove nucleotidi identificati correttamente con nessuna identificazione non rilevata)
- ▶ Lunghezze PolyT ≤ 10 (tutte le basi identificate in 819 ampliconi sequenziati contenenti una ripetizione PolyT di dieci nucleotidi identificati correttamente con nessuna identificazione non rilevata)
- ▶ Lunghezze PolyG ≤ 7 (tutte le basi identificate in 819 ampliconi sequenziati contenenti una ripetizione PolyG di sette nucleotidi identificati correttamente con una percentuale di identificazioni non rilevate di 1,0%)
- ▶ Lunghezze PolyC ≤ 6 (tutte le basi identificate in 2.457 ampliconi sequenziati contenenti una ripetizione PolyC di sei nucleotidi identificati correttamente con nessuna identificazione non rilevata)
- ▶ Lunghezze di dinucleotidi ripetuti $\leq 11x$ (tutte le basi identificate in 819 ampliconi sequenziati contenenti 11x di ripetizione di dinucleotidi sono state identificate correttamente con una percentuale di identificazioni non rilevate di 0,5%)
- ▶ Lunghezze di trinucleotidi ripetuti $\leq 5x$ (tutte le basi identificate in 819 ampliconi sequenziati contenenti 5x di ripetizione di trinucleotide sono state identificate correttamente con una percentuale di identificazioni non rilevate di 0,5%)
- ▶ Lunghezze inserzione ≤ 24 (66.343 delle 66.370 basi identificate in 819 ampliconi sequenziati contenenti un'inserzione di 24 nucleotidi identificate correttamente con una percentuale di identificazioni non rilevate di 1,2%; nessuna identificazione errata si è verificata nella regione contenente l'inserzione di 24 nucleotide)
- ▶ Lunghezze delezione ≤ 25 (tutte le basi identificate in 2.457 ampliconi sequenziati contenenti una delezione di 25 nucleotidi identificate correttamente con nessuna identificazione non rilevata)

Somatic

Lo studio qui descritto è stato condotto per valutare l'accuratezza dell'identificazione delle varianti del modulo Somatic Variant sullo strumento NextSeq 550Dx utilizzando NextSeq 550Dx High Output Reagent Kit v2 (300 cycles).

Questo studio ha utilizzato un saggio rappresentativo progettato per interrogare diversi geni che coprono 12.588 basi (150 ampliconi) su 23 diversi cromosomi. Nello studio, il DNA Platinum Genome è stato estratto da blocchi trattati in FFPE per generare sei campioni univoci da utilizzare per la valutazione.

Il campione di DNA GM12877 è stato diluito con il campione di DNA GM12878 per creare GM12877-D5 e GM12877-D7 come un set di varianti eterozigote univoche con frequenze della variante prossime al 5% e al 7%. Il campione di DNA GM12878 è stato diluito in modo simile con il campione di DNA GM12877 per creare GM12878-D5 e GM12878-D7. Ciascun campione è stato analizzato in triplicati fatta eccezione per i campioni diluiti, che sono stati analizzati in replicati di sei. È stato eseguito un totale di nove corse utilizzando tre strumenti di sequenziamento, tre lotti di reagenti e tre operatori su cinque giorni di avvio. L'accuratezza è stata

determinata per SNV, inserzioni e delezioni confrontando i risultati con il metodo di riferimento composito ben caratterizzato, Platinum Genomes versione 2016-1.0. Le regioni genomiche affidabili sono state definite in base a questo metodo di riferimento, se non altrimenti specificato.

Tabella 13 Riepilogo della concordanza per Somatic

Criteri	Osservazioni totali ¹	Risultato per osservazione ²	Risultato per corsa ³
PPA per SNV	378	98,9	99,9
PPA per le inserzioni	378	96,9	99,9
PPA per le delezioni	378	97,1	99,9
NPA	378	> 99,9	> 99,9
OPA	378	> 99,9	> 99,9

¹ Calcolato come il numero di campioni per corsa (42) x il numero di corse (9) = 378.

² Il valore più basso osservato per replicato di campione su tutte le nove corse.

³ Il valore più basso quando i dati di ciascuna corsa sono stati analizzati in aggregato.

La **Tabella 14** contiene i dati dello studio con la concordanza positiva e negativa espressa in percentuale in base ai singoli campioni, dove i risultati delle varianti sono confrontati con un metodo di riferimento composito ben caratterizzato per i calcoli della PPA. Sono stati combinati i tre tipi di varianti (SNV, inserzioni e delezioni). Poiché il metodo di riferimento fornisce solo risultati per le varianti di singolo nucleotide e inserzioni/delezioni, i risultati delle basi non varianti sono confrontati con la sequenza di riferimento del genoma umano versione hg19, per i calcoli della NPA.

Tabella 14 Concordanza per campione per Somatic

Campione	Percentuale di identificazione media	Previsto	TP	FN	Identificazioni varianti non rilevate	TN	FP	PPA	NPA	OPA
GM12877	98,7	2.052	2.025	0	27	318.682	15	100	> 99,9	> 99,9
GM12878	98,8	3.645	3.564	0	81	317.645	0	100	100	100
GM12879	99,8	2.592	2.538	0	54	323.614	2	100	> 99,9	> 99,9
GM12884	99,8	3.078	3.024	0	54	322.038	5	100	> 99,9	> 99,9
GM12885	99,8	3.294	3.213	0	81	322.121	0	100	100	100
GM12888	99,8	2.916	2.889	0	27	323.048	2	100	> 99,9	> 99,9
GM12877-D5	99,8	9.288	8.930	0	358	630.621	0	100	100	100
GM12877-D7	99,7	9.288	9.032	0	256	629.719	0	100	100	100
GM12878-D5	99,5	9.288	8.699	42	547	628.582	0	99,5	100	> 99,9
GM12878-D7	99,7	9.288	9.108	0	180	629.803	0	100	100	100

La **Tabella 15** contiene i dati dello studio in base ai singoli campioni, dove i risultati delle varianti sono confrontati con un metodo di riferimento composito ben caratterizzato. Il rilevamento è stato valutato per ciascun tipo di variante: SNV, inserzioni e delezioni, separatamente. Sono state escluse le posizioni dei riferimenti.

Tabella 15 Concordeza per campione in base al tipo di variante per Somatic

Campione	SNV			Inserzioni			Delezioni		
	Previsto	TP	FN	Previsto	TP	FN	Previsto	TP	FN
GM12877	999	999	0	567	567	0	486	459	0
GM12878	2.457	2.457	0	540	513	0	648	594	0
GM12879	1.539	1.539	0	567	540	0	486	459	0
GM12884	1.836	1.836	0	675	648	0	567	540	0
GM12885	2.025	2.025	0	675	648	0	594	540	0
GM12888	1.782	1.782	0	621	621	0	513	486	0
GM12877-D5	5.454	5.392	0	1.782	1.647	0	2.052	1.891	0
GM12877-D7	5.454	5.406	0	1.782	1.728	0	2.052	1.898	0
GM12878-D5	5.454	5.192	28	1.782	1.651	9	2.052	1.856	5
GM12878-D7	5.454	5.445	0	1.782	1.719	0	2.052	1.944	0

I dieci campioni sono stati ulteriormente analizzati per identificare piccole inserzioni e delezioni (Indel) (Tabella 16). Era presente un totale di 71 Indel con una dimensione compresa nell'intervallo 1–24 bp per le inserzioni e 1–25 bp per le delezioni.

Tabella 16 Riepilogo del rilevamento delle Indel per Somatic

Tipo di variante	Variante previste	TP	FN	Identificazioni varianti non rilevate	PPA
Inserzione	10.773	10.282	9	482	99,2
Delezione	11.502	10.667	5	830	> 99,9

I 150 ampliconi sono stati progettati per coprire diverso contenuto genomico. Il contenuto in GC degli ampliconi rientrava nell'intervallo compreso tra 0,19 e 0,87%. Gli ampliconi presentavano inoltre un intervallo di ripetizioni di singolo nucleotide (ad es., PolyA, PolyT), dinucleotide e trinucleotide. I dati sono stati compilati in base ai singoli ampliconi (Tabella 17) per determinare gli effetti del contenuto genomico sulla percentuale di identificazioni corrette. La percentuale di identificazioni corrette consiste di identificazioni delle varianti e dei riferimenti ed è inferiore al 100% se sono presenti identificazioni errate o identificazioni non rilevate.

Tabella 17 Accuratezza a livello di amplicone per Somatic

Amplicone	Cromo- soma	Avvio amplicone	Fine amplicone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplicone	Contenuto in GC	Identifi- cazioni corrette	Identifi- cazioni errate	Nessuna identifi- cazione	% di identifi- cazioni corrette
1	1	36450499	36450591	93	93	Indel	0,22	35.066	0	88	99,7
2	1	109465122	109465200	79	79	Poly A (5), Poly C (5), Indel	0,38	29.827	0	35	99,9
3	1	218353867	218353957	91	91	Indel	0,4	34.202	0	283	99,2
4	1	223906657	223906748	92	92	Indel	0,49	34.613	0	163	99,5
5	1	228526602	228526682	81	81	Poly G (5)	0,69	30.571	0	47	99,8
6	1	236372039	236372108	70	70	Poly T (10), Indel	0,39	26.452	0	8	100,0
7	1	247812041	247812128	88	88	Poly A (5), CT(3), TAA(3), Indel	0,27	33.148	0	116	99,7
8	2	55862774	55862863	90	90	Indel	0,28	33.928	0	92	99,7
9	2	87003930	87004009	80	80	Indel	0,38	30.218	0	22	99,9
10	2	177016721	177016805	85	81	N/A	0,65	30.616	0	2	> 99,9
11	2	186625727	186625801	75	75	Poly A (8)	0,35	28.017	0	499	98,3
12	2	190323504	190323591	88	88	Poly T (5)	0,42	33.207	0	57	99,8
13	2	200796740	200796826	87	87	Poly T (5), indel	0,31	32.524	9	718	97,8
14	2	212245049	212245139	91	91	Poly T (5), Poly A (6), Indel	0,3	33.972	0	456	98,7
15	2	228147052	228147144	93	93	N/A	0,43	35.051	0	103	99,7
16	2	235016350	235016422	73	73	Poly T (5), indel	0,42	27.459	0	136	99,5
17	3	4466229	4466321	93	93	AT(3), Indel	0,27	34.534	0	620	98,2
18	3	46620561	46620643	83	83	N/A	0,43	31.339	0	44	99,9
19	3	49851331	49851400	70	70	CT(3), Indel	0,49	26.373	0	87	99,7

Amplificone	Cromo- soma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identifi- cazioni corrette	Identifi- cazioni errate	Nessuna identifi- cazione	% di identifi- cazioni corrette
20	3	189713161	189713248	88	88	Poly A (5), Poly T (5), Poly A (9), TG (3)	0,41	32.829	0	857	97,5
21	3	190106030	190106104	75	74	Indel	0,57	27.925	0	47	99,8
22	4	2233667	2233744	78	78	Poly A (6)	0,26	29.327	4	162	99,4
23	4	7780541	7780637	97	97	Poly G (6), Poly T (5), Poly A (5)	0,42	36.585	0	117	99,7
24	4	15688604	15688681	78	78	N/A	0,29	29.427	0	57	99,8
25	4	56236521	56236586	66	62	Poly A (5), Indel	0,36	23.356	5	75	99,7
26	4	102839244	102839314	71	69	Poly A (5)	0,46	25.942	0	140	99,5
27	4	164446743	164446804	62	62	Poly A (7), Indel	0,27	22.944	0	560	97,6
28	5	1882081	1882158	78	75	N/A	0,78	28.299	0	53	99,8
29	5	14769061	14769144	84	84	GT(3), CCA(3)	0,62	31.658	0	94	99,7
30	5	41069808	41069871	64	64	N/A	0,39	24.120	0	72	99,7
31	5	74077114	74077196	83	83	Poly A (6), Indel	0,3	31.297	0	77	99,8
32	5	147475343	147475409	67	67	Poly T (5)	0,37	25.277	0	55	99,8
33	5	149323731	149323821	91	91	CT(4), AG(3)	0,55	34.308	0	90	99,7
34	5	155662213	155662287	75	75	Indel	0,43	28.266	0	163	99,4
35	6	6318713	6318814	102	102	Poly G (6)	0,68	38.489	0	67	99,8
36	6	24949983	24950074	92	92	Indel	0,63	34.730	0	46	99,9
37	6	31084900	31084999	100	94	GCT(5), Indel	0,61	35.057	0	483	98,6
38	6	32147987	32148084	98	98	Poly T (5), TCT(3), CTT(3)	0,55	36.647	0	406	98,9
39	6	32986864	32986958	95	95	Indel	0,53	35.681	0	238	99,3
40	6	33408498	33408583	86	86	Poly C (6)	0,7	32.438	0	70	99,8
41	6	41647401	41647495	95	94	Poly G (5), Indel	0,61	35.441	0	91	99,7
42	6	112435865	112435955	91	91	Poly A (5)	0,44	34.354	0	44	99,9

Amplificone	Cromo- soma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identifi- cazioni corrette	Identifi- cazioni errate	Nessuna identifi- cazione	% di identifi- cazioni corrette
43	7	22202076	22202148	73	73	N/A	0,44	27.575	0	28	99,9
44	7	66276100	66276187	88	88	Indel	0,35	33.060	0	213	99,4
45	7	77365735	77365821	87	87	Poly A (7), AG(4)	0,26	32.423	0	489	98,5
46	7	110939946	110940030	85	85	Indel	0,38	32.074	0	56	99,8
47	7	128533468	128533557	90	90	Poly G (5), Indel	0,62	33.791	0	281	99,2
48	7	149503875	149503965	91	91	Poly G (6), Poly C (6), Indel	0,71	34.316	0	82	99,8
49	7	154404519	154404599	81	66	N/A	0,31	24.901	0	47	99,8
50	7	156476507	156476599	93	93	Indel	0,35	35.067	0	87	99,8
51	8	1817312	1817394	83	83	N/A	0,42	31.365	0	9	> 99,9
52	8	24811020	24811109	90	89	Poly G (7), CTC(4), Indel	0,61	32.781	0	890	97,4
53	8	76518625	76518691	67	67	Indel	0,3	25.228	0	146	99,4
54	9	103054909	103055006	98	98	Poly G (6)	0,67	36.968	0	76	99,8
55	9	105586150	105586214	65	65	Indel	0,32	24.472	0	100	99,6
56	9	107620823	107620918	96	96	N/A	0,49	36.203	0	85	99,8
57	9	123769149	123769231	83	83	AT(3)	0,37	31.329	0	45	99,9
58	9	138995345	138995441	97	97	Poly C (6), Indel	0,68	36.472	0	201	99,5
59	10	5987120	5987198	79	78	Poly G (5), Indel	0,47	29.473	0	11	> 99,9
60	10	11784629	11784726	98	91	GC(3)	0,87	34.188	0	213	99,4
61	10	27317777	27317855	79	79	Poly T (5)	0,3	29.843	0	19	99,9
62	10	33018351	33018440	90	90	Poly A (5), Poly T (5)	0,2	33.968	0	68	99,8
63	10	45084159	45084253	95	95	Indel	0,35	35.829	0	81	99,8
64	10	55892599	55892687	89	88	AC(11), Indel	0,42	32.098	88	2.048	93,8
65	10	101611250	101611329	80	80	N/A	0,49	30.217	0	28	99,9
66	10	118351373	118351453	81	81	N/A	0,51	30.531	0	96	99,7

Amplificone	Cromo- soma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identifi- cazioni corrette	Identifi- cazioni errate	Nessuna identifi- cazione	% di identifi- cazioni corrette
67	11	8159816	8159912	97	96	N/A	0,45	36.105	0	192	99,5
68	11	30177648	30177717	70	70	Indel	0,46	26.318	0	153	99,4
69	11	47470345	47470444	100	100	N/A	0,65	37.785	0	24	99,9
70	11	59837679	59837740	62	62	Indel	0,37	23.368	0	68	99,7
71	11	64418856	64418957	102	102	N/A	0,59	38.546	0	10	> 99,9
72	11	93529612	93529684	73	73	Poly A (5)	0,4	27.516	0	78	99,7
73	11	101347052	101347136	85	85	N/A	0,42	32.083	0	48	99,9
74	11	102477336	102477426	91	91	Poly G (6)	0,55	34.047	0	369	98,9
75	11	118406285	118406369	85	85	Indel	0,53	32.065	0	74	99,8
76	11	120357801	120357885	85	85	Poly A (5), CA(3), Indel	0,34	32.083	0	47	99,9
77	11	125769313	125769397	85	85	GA(3)	0,52	32.103	0	27	99,9
78	12	2834770	2834853	84	84	Poly C (5), Indel	0,52	31.645	16	525	98,3
79	12	26811004	26811096	93	93	Poly A (7), AC(4)	0,33	34.824	0	330	99,1
80	12	30881766	30881846	81	81	N/A	0,49	30.497	0	121	99,6
81	12	88474105	88474175	71	71	Poly A (6)	0,35	26.773	0	65	99,8
82	12	120966872	120966966	95	95	Poly G (5)	0,68	35.830	9	72	99,8
83	13	24167504	24167576	73	73	N/A	0,52	27.498	0	114	99,6
84	13	25816961	25817049	89	88	Poly A (5), Poly T (7), Poly A (7), Indel	0,22	32.824	0	566	98,3
85	13	44880112	44880200	89	89	Indel	0,49	33.574	0	77	99,8
86	13	77665218	77665294	77	77	Indel	0,39	29.075	0	31	99,9
87	14	31619327	31619393	67	67	GA(3),TA(3)	0,39	25.313	0	13	99,9
88	14	39517884	39517966	83	83	N/A	0,25	31.360	0	22	99,9
89	14	46958962	46959034	73	72	Poly T (5), indel	0,19	26.499	0	717	97,4
90	14	58050030	58050110	81	81	Indel	0,38	30.494	0	133	99,6

Amplificone	Cromo- soma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identifi- cazioni corrette	Identifi- cazioni errate	Nessuna identifi- cazione	% di identifi- cazioni corrette
91	14	82390559	82390649	91	91	Indel	0,35	34.313	0	86	99,7
92	14	92549544	92549609	66	66	Poly A (5)	0,41	24.555	0	1.527	94,1
93	14	102808496	102808589	94	94	Indel	0,62	35.472	0	69	99,8
94	15	43170751	43170848	98	96	Poly C (5)	0,45	36.264	0	24	99,9
95	15	63446149	63446216	68	68	Indel	0,25	25.667	0	37	99,9
96	15	77879807	77879901	95	93	Poly G (5), Indel	0,68	34.745	0	432	98,8
97	15	81625334	81625428	95	95	Poly T (6)	0,43	35.870	0	40	99,9
98	15	85438263	85438334	72	71	Indel	0,65	26.762	0	76	99,7
99	15	89817413	89817503	91	91	N/A	0,36	34.286	0	112	99,7
100	15	89864274	89864343	70	70	Indel	0,56	26.449	0	11	> 99,9
101	16	1894910	1894972	63	63	N/A	0,27	23.809	0	5	> 99,9
102	16	28997904	28997998	95	95	Poly C (5)	0,67	35.860	0	50	99,9
103	16	53682908	53682994	87	87	TA(3)	0,41	32.835	0	60	99,8
104	16	57954406	57954509	104	104	Poly C (5)	0,67	39.177	0	144	99,6
105	16	85706375	85706465	91	91	Poly T (5), indel	0,37	34.075	0	323	99,1
106	17	3563920	3564008	89	89	GC(3)	0,64	33.632	0	11	> 99,9
107	17	3594191	3594277	87	87	Poly C (5), Indel	0,67	32.752	0	134	99,6
108	17	3970090	3970180	91	91	Indel	0,46	34.343	0	82	99,8
109	17	16084945	16085037	93	93	Indel	0,26	35.077	0	78	99,8
110	17	33998759	33998849	91	89	Poly T (5)	0,54	33.553	0	89	99,7
111	17	39589691	39589774	84	82	Poly A (13), Indel (x2)	0,29	30.554	53	2.296	92,9
112	17	41244394	41244484	91	91	Poly A (5)	0,34	34.360	0	38	99,9
113	17	45438866	45438957	92	92	Poly A (7), AT(3), AT(4), AT(4), Indel	0,26	34.367	0	418	98,8
114	17	61502432	61502510	79	79	Indel	0,41	29.751	0	119	99,6
115	17	64023582	64023667	86	86	Poly T (7)	0,22	32.176	0	340	99,0

Amplificone	Cromo- soma	Avvio amplificone	Fine amplificone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplificone	Contenuto in GC	Identifi- cazioni corrette	Identifi- cazioni errate	Nessuna identifi- cazione	% di identifi- cazioni corrette
116	17	72308237	72308320	84	84	GAG(3)	0,62	31.604	7	141	99,5
117	18	2616456	2616522	67	67	GA(3)	0,31	25.273	8	45	99,8
118	18	6980478	6980568	91	91	N/A	0,37	34.386	0	12	> 99,9
119	18	9888026	9888094	69	69	Poly A (6), TG(3)	0,43	25.692	0	399	98,5
120	18	38836999	38837073	75	75	Poly A (5), Indel	0,37	27.923	0	893	96,9
121	18	47405382	47405462	81	81	CTC(3), Indel	0,47	30.598	0	20	99,9
122	18	54815665	54815749	85	85	CT(3), Indel	0,45	31.969	0	161	99,5
123	18	59773996	59774060	65	65	N/A	0,48	24.531	0	48	99,8
124	19	625143	625241	99	99	N/A	0,59	37.298	0	124	99,7
125	19	18121418	18121491	74	74	N/A	0,68	27.881	0	109	99,6
126	19	18186574	18186643	70	70	N/A	0,64	26.442	0	26	99,9
127	20	746056	746149	94	94	N/A	0,61	35.501	0	31	99,9
128	20	10633195	10633276	82	82	AC(3)	0,59	30.951	0	72	99,8
129	20	17705633	17705708	76	76	CT(3)	0,58	28.686	0	42	99,9
130	20	21766821	21766890	70	70	GT(3),TG(4), Indel	0,46	26.372	0	88	99,7
131	20	25278421	25278521	101	101	Indel	0,63	38.159	0	20	99,9
132	20	50897302	50897368	67	67	Indel	0,36	25.188	0	544	97,9
133	20	62331904	62331994	91	88	Poly G (6)	0,73	32.969	0	309	99,1
134	20	62690860	62690946	87	87	Indel	0,57	32.818	0	77	99,8
135	21	30300823	30300888	66	66	Indel	0,35	24.758	9	181	99,2
136	21	33694176	33694273	98	98	Poly T (6), CA(3)	0,54	36.902	0	160	99,6
137	21	36710706	36710792	87	87	GT(3), Indel	0,39	32.841	0	48	99,9
138	21	46644924	46644992	69	69	Poly A (6), AG(3), Indel	0,32	25.939	0	280	98,9
139	21	46705575	46705664	90	90	Poly T (5), Poly A (6)	0,5	33.942	0	78	99,8
140	22	25750774	25750873	100	100	Indel	0,63	37.733	0	86	99,8

Amplicone	Cromo- soma	Avvio amplicone	Fine amplicone	Dimensione del frammento analizzato	Basi nella regione affidabile	Contenuto genomico dell'amplicone	Contenuto in GC	Identifi- cazioni corrette	Identifi- cazioni errate	Nessuna identifi- cazione	% di identifi- cazioni corrette
141	22	32439233	32439329	97	97	N/A	0,68	36.617	0	49	99,9
142	22	37409844	37409940	97	97	Indel	0,46	36.525	0	162	99,6
143	22	37637596	37637694	99	99	N/A	0,6	37.398	0	24	99,9
144	22	47081347	47081438	92	92	Indel	0,66	34.754	0	22	99,9
145	X	15870424	15870492	69	69	Poly T (5)	0,26	26.046	0	36	99,9
146	X	135288543	135288611	69	69	Poly C (5)	0,62	26.019	0	63	99,8
147	X	135290777	135290847	71	71	N/A	0,52	26.780	0	58	99,8
148	Sì	2655397	2655461	65	0	N/A	0,55	0	0	0	NA
149	Sì	2655519	2655609	91	0	N/A	0,48	0	0	0	NA
150	Sì	2655609	2655679	71	0	Poly A (5)	0,37	0	0	0	NA

I risultati del sequenziamento per il campione GM12878 sono stati confrontati con un genotipo altamente affidabile per NA12878, stabilito dal National Institutes of Standards and Technology (NIST) (v.2.19). Dei 150 ampliconi, 92 ampliconi sono rientrati completamente nelle regioni genomiche altamente affidabili, 41 ampliconi presentavano una sovrapposizione parziale e 17 ampliconi non presentavano alcuna sovrapposizione nella sequenza NIST. Questo ha fornito 10.000 coordinate per replicato per il confronto. Le identificazioni delle basi non varianti sono state confrontate con la sequenza di riferimento del genoma umano versione hg 19. I risultati dell'accuratezza sono mostrati nella [Tabella 18](#).

Tabella 18 Concordezza del campione GM12878 rispetto al database NIST per Somatic

Campione	N. di ampliconi	Percentuale di identificazione media	TP	FN	TN	FP	PPA	NPA	OPA
GM12878	133	98,8	2.808	0	258.488	0	100	100	100

In base ai dati forniti da questo studio Somatic di nove corse, lo strumento NextSeq 550Dx può sequenziare in modo coerente:

- ▶ Contenuto in GC $\geq 19\%$ (tutte le basi identificate in 378 ampliconi sequenziati con 19% di contenuto in GC identificato correttamente con una percentuale di identificazioni non rilevate di 2,6%)
- ▶ Contenuto in GC $\leq 87\%$ (tutte le basi identificate in 378 ampliconi sequenziati con 87% di contenuto in GC identificato correttamente con una percentuale di identificazioni non rilevate di 0,6%)
- ▶ Lunghezze PolyA ≤ 9 (tutte le basi identificate in 378 ampliconi sequenziati contenenti una ripetizione PolyA di nove nucleotidi identificati correttamente con una percentuale di identificazioni non rilevate di 2,5%)
- ▶ Lunghezze PolyT ≤ 10 (tutte le basi identificate in 378 ampliconi sequenziati contenenti una ripetizione PolyT di dieci nucleotidi identificati correttamente con una percentuale di identificazioni non rilevate inferiore a 0,1%)
- ▶ Lunghezze PolyG ≤ 6 (tutte le basi identificate in 2.268 ampliconi sequenziati contenenti una ripetizione PolyG di sei nucleotidi identificati correttamente con una percentuale di identificazioni non rilevate di 0,5%)
- ▶ Lunghezze PolyC ≤ 6 (tutte le basi identificate in 756 ampliconi sequenziati contenenti una ripetizione PolyC di sei nucleotidi identificati correttamente con una percentuale di identificazioni non rilevate di 0,4%)
- ▶ Lunghezze di dinucleotidi ripetuti $\leq 4x$ (tutte le basi identificate in 1.890 ampliconi sequenziati con 4x di ripetizione di dinucleotide sono state identificate correttamente con una percentuale di identificazioni non rilevate di 0,9%)
- ▶ Lunghezze di trinucleotidi ripetuti $\leq 5x$ (tutte le basi identificate in 378 ampliconi sequenziati contenenti 5x di ripetizione di trinucleotide sono state identificate correttamente con una percentuale di identificazioni non rilevate di 1,4%)
- ▶ Lunghezze inserzioni ≤ 23 (tutte le basi identificate in 378 ampliconi sequenziati contenenti una inserzione di 23 nucleotidi identificata correttamente con una percentuale di identificazioni non rilevate di 0,8%)
- ▶ Lunghezze delezione ≤ 25 (tutte le basi identificate in 1.134 ampliconi sequenziati contenenti una delezione di 25 nucleotidi identificata correttamente con una percentuale di identificazioni non rilevate di 0,7%)

Precisione

La precisione dello strumento NextSeq 550Dx è stata determinata analizzando 13 campioni univoci Platinum Genome utilizzando tre strumenti, tre lotti di reagenti e tre operatori per generare nove corse di sequenziamento nel corso di cinque giorni di avvio. Il saggio rappresentativo, i campioni e il metodo di riferimento sono gli stessi di quelli descritti per lo studio di accuratezza per Germline. I contributi di precisione sono stati determinati in base all'analisi del componente della variazione utilizzando VAF come variabile della risposta e calcolando le deviazioni standard a livello di componente per lo strumento, il lotto di reagenti, l'operatore e il giorno di avvio ([Tabella 19](#)). Il numero totale delle osservazioni utilizzate nell'analisi di ogni componente per la variabilità di strumento, operatore o lotto di reagenti era di 699, 176 e 235 per SNV, inserzioni e delezioni, rispettivamente.

Tabella 19 Risultati di precisione per lo strumento NextSeq 550Dx (deviazione standard)

Componente	Tipo di variante	Componente DS		Totale DS	
		Max	Mediana	Max	Mediana
Lotto	SNV	0,0076	0,0002	0,0833	0,0154
	Inserzione	0,0104	0,0000	0,0410	0,0157
	Delezione	0,0046	0,0005	0,0560	0,0187
Strumento	SNV	0,0114	0,0003	0,0840	0,0153
	Inserzione	0,0138	0,0009	0,0407	0,0161
	Delezione	0,0079	0,0008	0,0549	0,0187
Operatore	SNV	0,0226	0,0008	0,0841	0,0155
	Inserzione	0,0344	0,0010	0,0417	0,0164
	Delezione	0,0083	0,0013	0,0547	0,0187
Giorno	SNV	0,0277	0,0012	0,0825	0,0160
	Inserzione	0,0235	0,0012	0,0409	0,0169
	Delezione	0,0271	0,0014	0,0548	0,0188

Confronto dei metodi (piattaforma di sequenziamento)

Sono stati valutati campioni di sangue intero e in FFPE sullo strumento NextSeq 550Dx e sullo strumento MiSeqDx utilizzando TruSeq Custom Amplicon Kit Dx per i flussi di lavoro Germline e Somatic. La concordanza della frequenza delle varianti per i campioni di sangue e in FFPE è stata valutata utilizzando diversi saggi rappresentativi. La [Figura 2](#) rappresenta la correlazione VAF tra i due strumenti per un saggio rappresentativo e la [Tabella 20](#) riassume questa correlazione in base al pannello del saggio. Sulla base della forte correlazione tra lo strumento MiSeqDx e lo strumento NextSeq 550Dx, le caratteristiche prestazionali relative ai fattori pre-analitici (ad es., metodi di estrazione o sostanze interferenti) sono state determinate come applicabili a entrambi gli strumenti. Per ulteriori dettagli, fare riferimento all'insero della confezione di TruSeq Custom Amplicon Kit Dx.

Figura 2 Correlazione di VAF tra lo strumento MiSeqDx e lo strumento NextSeq 550Dx per campioni in FFPE (sinistra) e campioni di sangue (destra) utilizzando il saggio 1

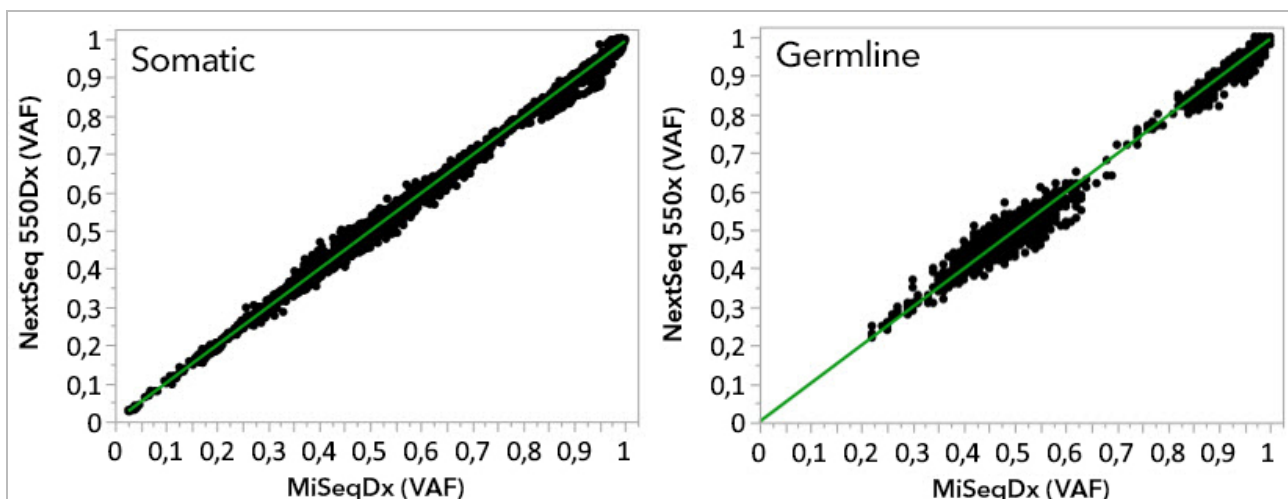


Tabella 20 Risultati del confronto del metodo utilizzando campioni univoci di sangue e in FFPE

Origine di gDNA	Saggio (pannello oligonucleotidi)	Replicati biologici (campioni)	Replicati tecnici (per campione)	Osservazioni (n. di varianti)	Pendenza	Intercetta	Correlazione (R ²)
Sangue	Saggio 1	45	2	8.369 ¹	0,992	0,002	0,995 ²
Sangue	Saggio 2	45	2	5.457	0,995	0,005	0,981
FFPE	Saggio 1	46	2	8.319	0,993	0,000	0,997 ²
FFPE	Saggio 3	40	1	280	0,969	0,015	0,978

¹ Due punti dei dati sono stati rimossi in base alla limitazione data per il modulo Germline Variant.

² Coefficiente di determinazione dei grafici VAF come illustrati nella Figura 2.

Riproducibilità

La riproducibilità dello strumento NextSeq 550Dx è stata valutata utilizzando i campioni Platinum Genome con un saggio rappresentativo progettato per interrogare diversi geni che coprono 12.588 basi su 23 diversi cromosomi utilizzando 150 ampliconi. L'analisi Germline consisteva di sette replicati di 13 campioni; l'analisi Somatic consisteva di sei replicati di sette campioni a diversi livelli VAF. I campioni sono stati preparati utilizzando TruSeq Custom Amplicon Kit Dx.

L'analisi è stata eseguita in tre sedi interne utilizzando un lotto NextSeq 550Dx High Output Reagent Kit v2 (300 cycles). In ciascun sito è stato utilizzato un solo strumento NextSeq 550Dx. In ciascun sito i test sono stati condotti da due operatori. Ogni operatore ha eseguito l'analisi in tre giorni di avvio non consecutivi per ciascun tipo di campione per un totale di 36 corse sui tre siti. Questo ha fornito 18 corse per ciascun flusso di lavoro per le linee germinali e per le varianti somatiche.

Germline

Le varianti Germline con livelli VAF di $\geq 0,2$ sono riportate come positive (variante). Per le varianti della linea germinale previste come positive, i dati sono stati valutati per percentuale di identificazioni non riuscite e percentuale di identificazioni positive corrette entro ciascun tipo di variante (SNV, inserzioni, delezioni). La **Tabella 21** riassume le percentuali osservate, assieme al livello superiore e inferiore di affidabilità (Lower Confidence Limit (LCL)/Upper Confidence Limit (UCL)) di 95% calcolato in base al metodo del punteggio di Wilson Score, per ciascun tipo di variante.

Tabella 21 Osservazioni delle identificazioni per i risultati positivi previsti in base al tipo di variante per Germline

Tipo di variante	Nessuna identificazione			Identificazione positiva corretta				
	Osservato	Totale	Percentuale	Osservato	Totale	Percentuale	LCL 95%	UCL 95%
SNV	16	110.376	0,014	110.349	110.360	99,99	99,98	99,99
Inserzioni	1026	37.044	2,77	36.018	36.018	100	99,99	100,00
Delezioni	648	34.776	1,86	34.128	34.128	100	99,99	100,00

Le varianti Germline con livelli VAF di $< 0,2$ sono riportate come negative (wild type). Per le posizioni delle linee germinali negative previste, i dati sono stati valutati per le percentuali di identificazioni non riuscite e identificazioni wild type. La **Tabella 22** riassume le percentuali osservate, assieme al livello superiore e inferiore di affidabilità (Lower Confidence Limit (LCL)/Upper Confidence Limit (UCL)) di 95% calcolato in base al metodo del punteggio di Wilson Score.

Tabella 22 Osservazione delle identificazioni per i risultati negativi previsti per Germline

Tipo di variante	Nessuna identificazione			Identificazione negativa corretta				
	Osservato	Totale	Percentuale	Osservato	Totale	Percentuale	LCL 95%	UCL 95%
Wild type	4.883	19.600.182	0,025	19.595.299	19.595.299	100	100,00	100,00

Le varianti Germline con un livello VAF di $\geq 0,2$ e $< 0,7$ sono identificate come eterozigote positive per la variante e le varianti con un livello VAF di $\geq 0,7$ sono identificate come omozigote positive per la variante.

I campioni Germline con varianti eterozigote sono stati utilizzati per determinare se la variabilità inerente del saggio incide sull'identificazione del genotipo. Per questo, è stato determinato il valore Cx per entrambi i cutoff (0,2 per i genotipi eterozigoti e 0,7 per i genotipi omozigoti), dove x rappresenta la proporzione di analisi ripetute che hanno superato il valore di cutoff. In relazione al cutoff inferiore di 0,2 VAF, il valore Cx era $\geq 99,999\%$ indicando quindi che $\geq 99,999\%$ di varianti eterozigote verranno identificate come eterozigote.

In relazione al cutoff superiore di 0,7 VAF, il valore Cx era $\leq 0,001\%$ indicando quindi che $\leq 0,001\%$ di varianti eterozigote verranno identificate come omozigote. La **Tabella 23** riassume i risultati in base al tipo di variante.

Le varianti Germline con un livello VAF di $\geq 0,2$ e $< 0,7$ sono identificate come eterozigote positive per la variante e le varianti con un livello VAF di $\geq 0,7$ sono identificate come omozigote positive per la variante. I campioni Germline con varianti eterozigote sono stati utilizzati per determinare se la variabilità inerente del saggio incide sull'identificazione del genotipo. Per questo, è stato determinato il valore Cx per entrambi i cutoff (0,2 per i genotipi eterozigoti e 0,7 per i genotipi omozigoti), dove x rappresenta la proporzione di analisi ripetute che hanno superato il valore di cutoff. In relazione al cutoff inferiore di 0,2 VAF, il valore Cx era $\geq 99,999\%$ indicando quindi che $\geq 99,999\%$ di varianti eterozigote verranno identificate come eterozigote. Per il cutoff superiore di 0,7 VAF, il valore Cx era $\leq 0,001\%$ indicando quindi che $\leq 0,001\%$ di varianti eterozigote verranno identificate come omozigote. La **Tabella 23** riassume i risultati in base al tipo di variante.

Tabella 23 Valori Cx Germline per le varianti eterozigote

Tipo di variante	Cutoff a 0,2 VAF	Cutoff a 0,7 VAF
	$\geq C99,999\%$	$\leq C0,001\%$
SNV	94/94	94/94
Inserzioni	24/24	24/24
Delezioni	35/35	35/35
Totale	153	153

Somatic

Le varianti Somatic con livelli VAF di $\geq 0,026$ sono riportate come positive (variante). Inoltre, per questa analisi sono state considerate equivoche le osservazioni con livelli VAF $\geq 0,01$ e $< 0,026$ (né positive né negative, indicate come frequenza della variante bassa). Per valutare le prestazioni, i risultati sono stati calcolati in tre modi:

- ▶ Caso migliore: qualsiasi risultato equivoco è stato considerato un'identificazione positiva corretta (in accordo con i risultati previsti)
- ▶ Caso peggiore: qualsiasi risultato equivoco è stato considerato un'identificazione errata (in disaccordo con i risultati previsti)
- ▶ Caso di esclusione: qualsiasi risultato equivoco è stato escluso dall'analisi

Tre tabelle, **Tabella 24**, **Tabella 25** e **Tabella 26**, riassume i risultati delle identificazioni per il caso migliore, il caso peggiore e il caso di esclusione, rispettivamente, assieme ai livelli di affidabilità inferiore e superiore al 95% (LCL/UCL) calcolati utilizzando il metodo del punteggio di Wilson.

Tabella 24 Osservazioni delle identificazioni per i risultati positivi previsti in base al tipo di variante (caso migliore) per Somatic

Tipo di variante	Identificazione positiva corretta				
	Osservato	Totale	Percentuale	LCL 95%	UCL 95%
SNV	54.346	54.346	100	99,99	100,00
Inserzioni	18.036	18.036	100	99,98	100,00
Delezioni	18.381	18.381	100	99,98	100,00

Tabella 25 Osservazioni delle identificazioni per i risultati positivi previsti in base al tipo di variante (caso peggiore) per Somatic

Tipo di variante	Identificazione positiva corretta				
	Osservato	Totale	Percentuale	LCL 95%	UCL 95%
SNV	54.346	54.346	100	99,99	100,00
Inserzioni	18.000	18.036	99,8	99,72	99,86
Delezioni	18.381	18.381	100	99,98	100,00

Tabella 26 Osservazioni delle identificazioni per i risultati positivi previsti in base al tipo di variante (casi equivoci rimossi) per Somatic

Tipo di variante	Identificazione positiva corretta				
	Osservato	Totale	Percentuale	LCL 95%	UCL 95%
SNV	54.346	54.346	100	99,99	100,00
Inserzioni	18.000	18.000	100	99,98	100,00
Delezioni	18.381	18.381	100	99,98	100,00

Le varianti Somatic con livelli VAF di $< 0,01$ sono riportate come identificazioni negative (wild type). Per le posizioni delle varianti somatiche negative previste, i dati sono stati valutati per la percentuale di identificazioni non rilevate e la percentuale di identificazioni wild type. Le identificazioni wild type sono state determinate escludendo le identificazioni non rilevate e sottraendo dal totale le identificazioni osservate che rientravano nella zona equivoca (livelli VAF di $\geq 0,01$ e $< 0,026$) nonché le identificazioni errate che risultavano al di sopra del livello di cutoff (livelli VAF di $\geq 0,026$). La [Tabella 27](#) riassume i risultati osservati, totali e in percentuale delle posizioni somatiche negative per la percentuale di identificazioni non rilevate e la percentuale di identificazioni wild type corrette con livelli di affidabilità inferiore e superiore al 95% (LCL/UCL) calcolati in base al metodo del punteggio di Wilson.

Tabella 27 Osservazione delle identificazioni per i risultati negativi previsti per Somatic

Tipo di variante	Nessuna identificazione			Identificazione corretta						
	Osservato	Totale	Percentuale	Equivoca	Errato	Corretto	Totale	Percentuale	LCL 95%	UCL 95%
Wild type	36.326	8.909.676	0,408	2254	121	8.870.975	8.873.350	99,97	99,972	99,974

Sono stati valutati i campioni per Somatic a diversi livelli VAF per la stessa variante per determinare C95 del saggio (entro ciascun tipo di variante). Per valutare la variabilità in prossimità del cutoff del saggio, sono stati utilizzati i campioni i cui livelli di VAF previsti rientravano nell'intervallo tra 0,02 e 0,07. C95 è stato determinato per ciascuna variante, con il valore più alto di C95 per ciascun tipo di variante riportato nella [Tabella 28](#).

Tabella 28 Riepilogo C95 per Somatic

Tipo di variante	N.	C95
SNV	74	0,0613
Inserzione	24	0,0573
Delezione	33	0,0575

NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles)

Descrizione generale

NextSeq 550Dx è supportato da due kit di reagenti: NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Per dimostrare che NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) è in grado di soddisfare i requisiti di prestazione analitica verificati e convalidati con NextSeq 550Dx High Output Reagent Kit v2 (300 cycles), sono stati condotti studi specifici con NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Sono state eseguite due preparazioni di librerie utilizzando TruSeq Custom Amplicon Kit Dx: la prima con il flusso di lavoro Germline e la seconda con il flusso di lavoro Somatic. Le librerie di ciascun flusso di lavoro sono state testate con tre lotti di NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) utilizzando tre strumenti NextSeq 550Dx. Inoltre, nei test di ogni flusso di lavoro è stata inclusa una corsa singola con NextSeq 550Dx High Output Reagent Kit v2 (300 cycles).

Sensibilità analitica: limite del campione bianco (Limit of Blank, LoB) e limite del rilevamento (Limit of Detection, LoD)

La verifica con NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) ha dimostrato che lo strumento NextSeq 550Dx potrebbe rilevare varianti a 0,05 VAF con un errore di tipo II $\leq 0,05$ e che il cutoff VAF di 0,026 utilizzato dal modulo Somatic Variant (con LoB efficace) supporta un errore di tipo I $\leq 0,01$. Sulla base di queste affermazioni ci si aspetta che una variante a 0,05 VAF sia maggiore o uguale a 0,026 VAF il 95% delle volte e che una posizione wild type sia inferiore a 0,026 VAF il 99% delle volte. Per assicurarsi che queste affermazioni siano soddisfatte con NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles), sono state condotte misurazioni ripetute sullo strumento NextSeq 550Dx con campioni wild type (campioni LoB) e con campioni contenenti varianti a 0,05 VAF (campioni LoD) utilizzando NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). È stato quindi effettuato un confronto tra la proporzione di identificazioni al di sopra e al di sotto del cutoff di 0,026 e le affermazioni stabilite con NextSeq 550Dx High Output Reagent Kit v2 (300 cycles).

I test includevano due campioni LoD ciascuno con un unico set di varianti mirate a 0,05 VAF e campioni corrispondenti LoB wild type per le varianti obiettivo del test. Per la preparazione delle librerie, i campioni LoD e LoB sono stati elaborati in replicati di otto e sette, rispettivamente, utilizzando TruSeq Custom Amplicon Kit Dx. Le librerie sono state inizialmente sequenziate utilizzando NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) per identificare le varianti/le coordinate genomiche per la valutazione LoB/LoD con NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Tutte le varianti con VAF medio compreso tra 0,045 e 0,055 (varianti LoD) in base ai risultati di NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) sono state utilizzate per l'analisi LoD (N = 51 varianti). Per l'analisi LoB, sono state valutate le 51 coordinate genomiche corrispondenti. Per la valutazione di NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles), le librerie sono state sequenziate in tre corse in tre giorni consecutivi utilizzando lo stesso strumento e lo stesso lotto di kit di reagenti. Il test comprendeva 24 replicati per ciascuna delle 51 varianti LoD e 21 replicati per ciascuna delle corrispondenti posizioni wild type. La proporzione di identificazioni wild type con VAF $< 0,026$ è riportata nella [Tabella 29](#). La proporzione di identificazioni di varianti LoD con VAF maggiore o uguale a 0,026 sono riportate nella [Tabella 30](#).

Tabella 29 Proporzioni di identificazioni $< 0,026$ per le posizioni wild type (valutazione delle affermazioni LoB)

Tipo di variante	Posizioni valutate	Osservazioni totali	N. di misurazioni VAF $\geq 2,6\%$	Proporzione $< 2,6\%$	Proporzione 95% Intervallo di confidenza
SNV	32	672	0	1	0,994 – 1
Inserzione	11	231	0	1	0,984 – 1
Delezione	8	168	0	1	0,978 – 1

Tabella 30 Proporzione di identificazioni $\geq 0,026$ VAF per le varianti LoD (valutazione delle identificazioni LoD)

Tipo di variante	Posizioni valutate	Osservazioni totali	N. di misurazioni VAF < 2,6%	N. di misurazioni VAF $\geq 2,6\%$	Proporzione $\geq 2,6\%$	Proporzione 95% Intervallo di confidenza
SNV	32	768	1	767	0,999	0,993 – 1
Inserzione	11	264	3	261	0,989	0,967 – 0,996
Delezione	8	192	2	190	0,99	0,963 – 0,997

Accuratezza

Germline

Il seguente studio è stato condotto per valutare l'accuratezza di identificazione delle varianti del modulo Germline Variant sullo strumento utilizzando NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Dodici campioni Platinum Genome univoci sono stati analizzati utilizzando un saggio rappresentativo. È stato eseguito un totale di 11 corse utilizzando tre strumenti NextSeq 550Dx e tre lotti NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).

L'accuratezza è stata determinata per SNV, inserzioni e delezioni confrontando i risultati sul metodo di riferimento composito ben caratterizzato, Platinum Genomes versione 2016-1.0. Vengono riportati come riferimento i risultati di accuratezza dalla corsa di sequenziamento singola ottenuti utilizzando NextSeq 550Dx High Output Reagent Kit v2 (300 cycles). Un riepilogo dei risultati è riportato in [Tabella 31](#).

Tabella 31 Riepilogo della concordanza per Germline

Criteri	Osservazioni totali (v2.5) ¹	Risultato per osservazione (v2.5) ²	Risultato per osservazione (v2) ³	Risultato per corsa (v2.5) ⁴	Risultato per corsa (v2) ⁴
PPA per SNV	1.056	98,7	98,7	> 99,9	> 99,9
PPA per le inserzioni	1.056	100	100	100	98,9
PPA per le delezioni	1.056	95,2	95,2	> 99,9	100
NPA	1.056	100	100	100	100
OPA	1.056	> 99,9	> 99,9	> 99,9	> 99,9

¹Calcolato come numero di campioni per corsa x numero di corse (96 campioni per corsa x 11 corse = 1.056 osservazioni).

²Valore più basso osservato per campione replicato in tutte le corse (sulla base di 11 corse con NextSeq 550Dx High Output Reagent Kit v2.5).

³Valore più basso osservato per campione replicato in 1 corsa (96 osservazioni totali).

⁴ Valore più basso quando i dati di ciascuna corsa vengono analizzati in aggregato.

Somatic

Il seguente studio è stato condotto per valutare l'accuratezza di identificazione delle varianti del modulo Somatic Variant sullo strumento NextSeq 550Dx utilizzando NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles). Sono stati testati dieci campioni Platinum Genome FFPE (due con varianti diluite fino a 0,05 VAF) utilizzando un saggio rappresentativo. È stato eseguito un totale di 11 corse utilizzando tre strumenti NextSeq 550Dx e tre lotti di NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).

L'accuratezza è stata determinata per SNV, inserzioni e delezioni confrontando i risultati sul metodo di riferimento composito ben caratterizzato, Platinum Genomes versione 2016-1.0. Vengono riportati come riferimento i risultati di accuratezza dalla corsa di sequenziamento singola ottenuti utilizzando NextSeq 550Dx High Output Reagent Kit v2 (300 cycles). Un riepilogo dei risultati è riportato in [Tabella 32](#).

Tabella 32 Riepilogo della concordanza per Somatic

Criteri	Osservazioni totali (v2.5) ¹	Risultato per osservazione (v2.5) ²	Risultato per osservazione (v2) ³	Risultato per corsa (v2.5) ⁴	Risultato per corsa (v2) ⁴
PPA per SNV	528	100	100	100	100
PPA per le inserzioni	528	96,9	96,9	> 99,9	> 99,9
PPA per le delezioni	528	100	100	100	100
NPA	528	> 99,9	> 99,9	> 99,9	> 99,9
OPA	528	> 99,9	> 99,9	> 99,9	> 99,9

¹Calcolato come numero di campioni per corsa x numero di corse (48 campioni per corsa x 11 corse = 528 osservazioni).

²Valore più basso osservato per campione replicato in tutte le corse (sulla base di 11 corse con NextSeq 550Dx High Output Reagent Kit v2.5).

³Valore più basso osservato per campione replicato in 1 corsa (96 osservazioni totali).

⁴ Il valore più basso quando i dati di ciascuna corsa sono stati analizzati in aggregato.

Precisione

Germline

La precisione di NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) con il modulo Germline Variant è stata valutata utilizzando campioni Platinum Genome e un saggio rappresentativo. Il test effettuato consisteva in una singola preparazione della libreria utilizzando TruSeq Custom Amplicon Kit Dx e comprendeva 12 campioni elaborati con otto replicati ciascuno. Le librerie sono state sequenziate con tre lotti di NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) e tre strumenti NextSeq 550Dx per un totale di nove corse di sequenziamento.

I campioni Germline con varianti eterozigote sono stati utilizzati per determinare se la variabilità inerente del saggio incide sull'identificazione del genotipo (N = 153 varianti eterozigote uniche). Per questo, è stato determinato il valore Cx per entrambi i cutoff del modulo Germline Variant (0,2 per i genotipi eterozigoti e 0,7 per i genotipi omozigoti), dove x rappresenta la proporzione di analisi ripetute che hanno superato il valore di cutoff. Per il cutoff inferiore di 0,2 VAF, la variante con il Cx minimo per NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) è risultata essere > 99,9%, indicando che un valore > 99,9% delle varianti eterozigote viene identificato come eterozigote. Per il cutoff superiore di 0,7 VAF, la variante con il Cx massimo per NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) è risultata essere < 1,5%, indicando che un valore ≤ 1,5% delle varianti eterozigote viene identificato come omozigote. La [Tabella 33](#) riepiloga i risultati in base al tipo di variante. Vengono riportati come riferimento i valori Cx dalla corsa di sequenziamento singola ottenuti utilizzando NextSeq 550Dx High Output Reagent Kit v2 (300 cycles).

Tabella 33 Valori Cx Germline per le varianti eterozigote

Tipo di variante	N.	Cutoff a 0,2 VAF		Cutoff a 0,7 VAF	
		Min Cx (v2.5) ¹	Min Cx (v2) ²	Max Cx (v2.5) ¹	Max Cx (v2) ²
SNV	94	> 99,9%	> 99,9%	1,5%	1,0%
Inserzioni	24	100%	100%	0%	< 0,1%
Delezioni	35	100%	> 99,9%	< 0,1%	< 0,1%

¹Valori Cx basati sulle stime della deviazione standard totale dall'analisi delle componenti della varianza.

²Valori Cx basati sulle deviazioni standard del campione.

Somatic

La precisione di NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) con il modulo Somatic Variant è stata valutata utilizzando campioni Platinum Genome FFPE e un saggio rappresentativo. Il test effettuato consisteva in una singola preparazione della libreria utilizzando TruSeq Custom Amplicon Kit Dx e comprendeva due campioni con otto replicati ciascuno. Le librerie sono state sequenziate utilizzando tre lotti di NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) e tre strumenti NextSeq 550Dx per un totale di nove corse di sequenziamento.

Le varianti somatiche con livelli VAF previsti $\leq 0,10$ VAF (N = 131 varianti uniche) sono state utilizzate per valutare la variabilità dello strumento nelle vicinanze del cutoff VAF del modulo Somatic Variant (le varianti somatiche con livello VAF $\geq 0,026$ vengono indicate come positive alla variante). Per ciascuna delle varianti somatiche sono stati determinati i valori C95. I valori C95 rappresentano la VAF in corrispondenza della quale la probabilità di un valore superiore rispetto al cutoff VAF del modulo Somatic Variant è del 95%. Nella **Tabella 34** sono riportati i valori C95 più alti per tipo di variante. Vengono riportati come riferimento i risultati C95 dalla corsa di sequenziamento singola ottenuti utilizzando NextSeq 550Dx High Output Reagent Kit v2 (300 cycles).

Tabella 34 Riepilogo C95 per Somatic

Tipo di variante	N. di varianti valutate	C95 (v2.5) ¹	C95 (v2) ²
SNV	74	0,064	0,063
Inserzioni	24	0,062	0,061
Delezioni	33	0,060	0,060

¹Valori C95 basati sulle stime della deviazione standard totale dall'analisi delle componenti della varianza.

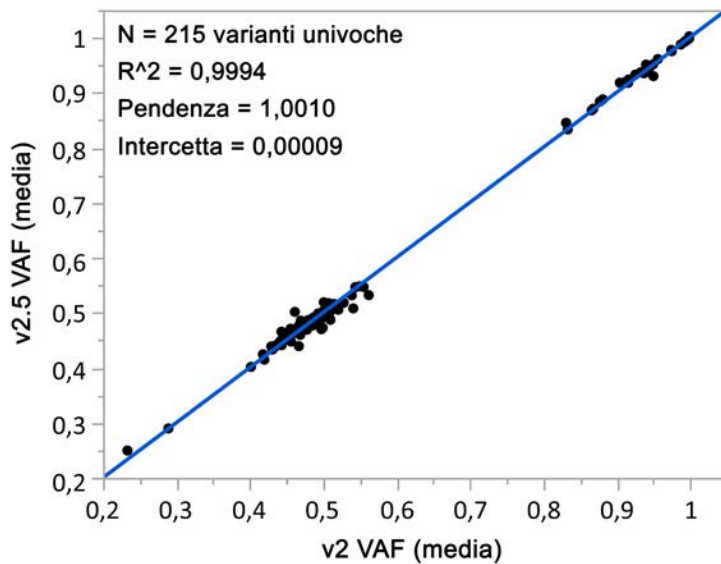
²Valori C95 basati sulle deviazioni standard del campione.

Confronto dei metodi (kit di reagenti)

Germline

Le frequenze VAF medie di 215 varianti uniche sono state valutate con il kit NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e con il kit NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) utilizzando i risultati generati dal modulo Germline Variant. Le medie VAF sono state calcolate a partire da 11 corse di sequenziamento (v2.5) e una corsa di sequenziamento (v2). Sono stati utilizzati almeno otto replicati per calcolare la media per ogni variante. La **Figura 3** mostra la correlazione VAF tra i due kit di reagenti. Sulla base della forte correlazione VAF lineare e della similitudine dei risultati tra i kit di reagenti, è stato determinato che le caratteristiche delle prestazioni inizialmente verificate e convalidate con NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) con il modulo Germline Variant possono essere applicate a NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).

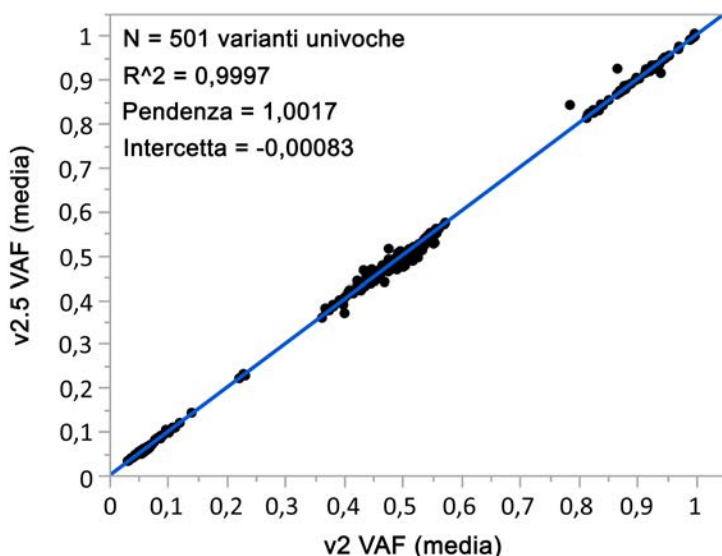
Figura 3 Correlazione della frequenza delle varianti alleliche (VAF) del modulo Germline Variant tra NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).



Somatic

Le frequenze VAF medie di 501 varianti uniche sono state valutate con il kit NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e con il kit NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles) utilizzando i risultati generati dal modulo Somatic Variant. Le medie VAF sono state calcolate a partire da 11 corse di sequenziamento (v2.5) e una corsa di sequenziamento (v2). Sono stati utilizzati almeno tre replicati per calcolare la media per ogni variante unica. La [Figura 4](#) mostra la correlazione VAF tra i due kit di reagenti. Sulla base della correlazione VAF e della similitudine dei risultati tra i kit di reagenti, è stato determinato che le caratteristiche delle prestazioni verificate e convalidate con NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) con il modulo Somatic Variant possono essere applicate a NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).

Figura 4 Correlazione della frequenza delle varianti alleliche (VAF) del modulo Somatic Variant tra NextSeq 550Dx High Output Reagent Kit v2 (300 cycles) e NextSeq 550Dx High Output Reagent Kit v2.5 (300 cycles).



Cronologia revisioni

Documento	Data	Descrizione della modifica
Documento n. 1000000030326 v06	Maggio 2022	Aggiornamenti apportati per correggere contenuti inseriti erroneamente dal software di redazione.
Documento n. 1000000030326 v05	Novembre 2021	Aggiunta la dichiarazione "Avvertenze e precauzioni" alla segnalazione di incidenti gravi. Aggiunta una dichiarazione ai "Principi della procedura" che specifica l'utente previsto. Rimosso il riferimento a High Output Reagent Kit v2 (300 cycles). Aggiunto il riferimento a "High Output Reagent Kit v2.5 (75 cycles)".
Documento n. 1000000030326 v04	Agosto 2021	Aggiunta la tabella Cronologia revisioni. Aggiornati gli indirizzi dei rappresentanti autorizzati nell'Unione Europea.

Brevetti e marchi di fabbrica

Questo documento e il suo contenuto sono di proprietà di Illumina, Inc. e delle aziende ad essa affiliate ("Illumina") e sono destinati esclusivamente ad uso contrattuale da parte dei clienti di Illumina, per quanto concerne l'utilizzo dei prodotti qui descritti, con esclusione di qualsiasi altro scopo. Questo documento e il suo contenuto non possono essere usati o distribuiti per altri scopi e/o in altro modo diffusi, resi pubblici o riprodotti, senza previa approvazione scritta da parte di Illumina. Mediante questo documento, Illumina non trasferisce a terzi alcuna licenza ai sensi dei suoi brevetti, marchi, copyright, o diritti riconosciuti dal diritto consuetudinario, né diritti simili di alcun genere.

Al fine di assicurare un uso sicuro e corretto dei prodotti qui descritti, le istruzioni riportate in questo documento devono essere scrupolosamente ed esplicitamente seguite da personale qualificato e adeguatamente formato. Leggere e comprendere a fondo tutto il contenuto di questo documento prima di usare tali prodotti.

LA LETTURA INCOMPLETA DEL CONTENUTO DEL PRESENTE DOCUMENTO E IL MANCATO RISPETTO DI TUTTE LE ISTRUZIONI IVI CONTENUTE POSSONO CAUSARE DANNI AL/I PRODOTTO/I, LESIONI PERSONALI A UTENTI E TERZI E DANNI MATERIALI E RENDERANNO NULLA QUALSIASI GARANZIA APPLICABILE AL/I PRODOTTO/I.

ILLUMINA NON SI ASSUME ALCUNA RESPONSABILITÀ DERIVANTE DALL'USO IMPROPRIO DEL/DEI PRODOTTO/I QUI DESCRITTI (INCLUSI SOFTWARE O PARTI DI ESSO).

© 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.

Informazioni di contatto



Illumina
5200 Illumina Way
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (fuori dal Nord America)
techsupport@illumina.com
www.illumina.com



Illumina Netherlands B.V.
Steenoven 19
5626 DK Eindhoven
Paesi Bassi

Sponsor Australiano
Illumina Australia Pty Ltd
Nursing Association Building
Level 3, 535 Elizabeth Street
Melbourne, VIC 3000
Australia

Etichettatura del prodotto

Per un riferimento completo dei simboli che si trovano sulla confezione del prodotto e sull'etichettatura, fare riferimento alla legenda dei simboli per il kit in uso alla pagina Web support.illumina.com.