

Insero della confezione

PER USO DIAGNOSTICO IN VITRO
SOLO PER ESPORTAZIONE

Uso previsto

Strumento NovaSeq 6000Dx è destinato al sequenziamento di librerie di DNA quando utilizzato con test per diagnostica *in vitro* (IVD). Strumento NovaSeq 6000Dx è destinato all'uso con specifici reagenti IVD e software analitici registrati, certificati o approvati.

Principi della procedura

Illumina® Strumento NovaSeq 6000Dx è destinato al sequenziamento di librerie di DNA con test diagnostici *in vitro*. Come input, NovaSeq 6000Dx 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 (Sequencing By Synthesis, SBS). 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 su Illumina DRAGEN Server per NovaSeq 6000Dx incluso e necessario per elaborare le identificazioni delle basi. Per l'analisi secondaria NovaSeq 6000Dx utilizza due diverse applicazioni in base al flusso di lavoro. Per l'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina, 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 e gVCF contengono informazioni sulle varianti germinali o somatiche (a seconda del flusso di lavoro selezionato) trovate in posizioni specifiche in un genoma di riferimento.

Duplici modalità di funzionamento

NovaSeq 6000Dx include un singolo disco rigido di avvio con modalità separate per la diagnostica *in vitro* (In Vitro Diagnostic, IVD) e per il solo uso di ricerca (Research Use Only, RUO). La modalità viene selezionata mediante un pulsante nella schermata Sequencing (Sequenziamento). La modalità selezionata è chiaramente indicata nell'interfaccia su tutte le schermate. I test di sequenziamento IVD, compresa l'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina nei flussi di lavoro della linea germinale e somatico, vengono eseguiti in modalità IVD. I reagenti per il sequenziamento per IVD possono essere utilizzati solo in modalità IVD. Le caratteristiche prestazionali e i limiti della procedura per NovaSeq 6000Dx sono stati stabiliti utilizzando l'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina in modalità IVD.

Limiti della procedura

1. Solo per uso diagnostico *in vitro*.
2. L'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina, quando utilizzata con Kit reagenti NovaSeq 6000Dx S2 v1.5 (300 cicli) e Kit Reagenti NovaSeq 6000Dx S4 v1.5 (300 cicli) è in grado di fornire:
 - Output di sequenziamento:
 - $\geq 1,0$ terabasi (TB) con il kit S2
 - $\geq 3,0$ TB con il kit S4
 - Lunghezza di lettura (in corse paired-end) da 2 x 150 coppie di basi (base pairs, bp).
 - Basi superiori a Q30 $\geq 85\%$ alla lunghezza della lettura da 2 x 150 bp. L'85% o più delle identificazioni delle basi ha un punteggio qualitativo sulla scala Phred superiore a 30, il che indica un'accuratezza dell'identificazione delle basi è superiore al 99,9%.
3. Le inserzioni di lunghezza > 18 bp e le delezioni di lunghezza > 21 bp non sono state convalidate.
4. 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.
5. Le MNV piccole sono riportate come varianti separate nel file VCF di output.
6. 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.
7. Limitazioni specifiche per il modulo Germline:
 - NovaSeq 6000Dx che utilizza il flusso di lavoro di analisi Germline della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina è progettato per fornire risultati qualitativi per l'identificazione di varianti germinali (ad esempio, omozigote, eterozigote, wild type).
 - La variazione del numero di copie può incidere sulla possibilità che una variante venga identificata come omozigote o eterozigote.
 - Il sistema non segnalerà più di due varianti in un singolo locus, anche in presenza di variazioni del numero di copie.
8. Limitazioni specifiche per il modulo Somatic:
 - NovaSeq 6000Dx che utilizza il flusso di lavoro di analisi Somatic della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina è progettato per fornire risultati qualitativi per l'identificazione di varianti somatiche (ad esempio, presenza di una variante somatica).
 - Il flusso di lavoro per l'analisi Somatic della generazione di file FASTQ e VCF non è in grado di distinguere tra varianti germinali e somatiche. Il flusso di lavoro 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 germinale.

- 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.
- Se viene identificato più di un allele di variante nello stesso locus, nessuno degli alleli verrà segnalato come variante di passaggio. Invece, verrà riportato l'intero set di alleli, ma filtrato tramite il tag multiallelico.

Procedure di controllo qualità

Il software NovaSeq 6000Dx valuta ciascuna corsa, campione e identificazione delle basi rispetto a metriche di controllo qualità. Nella preparazione delle librerie sono raccomandati controlli positivi e negativi e questi devono essere valutati. Valutare i controlli nel modo seguente.

- Controllo negativo (nessun controllo del modello) 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.

Componenti del prodotto

Illumina NovaSeq 6000Dx è composto da:

1. Strumento NovaSeq 6000Dx (N. di catalogo 20068232)
2. I componenti software di Strumento NovaSeq 6000Dx includono:

Applicazione software	Sede di installazione	Funzione	Descrizione
Software operativo NovaSeq	NovaSeq 6000Dx	Controlla il funzionamento dello strumento	Software operativo NovaSeq (NVOS) gestisce il funzionamento dello strumento durante il sequenziamento e genera immagini da usare con il software Real-Time Analysis (RTA).

Applicazione software	Sede di installazione	Funzione	Descrizione
Software Real-Time Analysis (RTA)	NovaSeq 6000Dx	Esegue l'analisi primaria	L'applicazione software RTA converte le immagini generate da NVOS per ciascuna tile per ciclo della corsa di sequenziamento in file di identificazione delle basi. I file di identificazione delle basi sono gli input per i moduli dell'applicazione su Illumina DRAGEN Server per NovaSeq 6000Dx. L'applicazione software RTA non include un'interfaccia utente.
Illumina Run Manager	Illumina DRAGEN Server	Impostazione e gestione dei controlli della corsa	Illumina Run Manager fornisce la gestione degli utenti e degli strumenti, ospita il software dell'applicazione e consente l'uso dei moduli di analisi secondaria di genomica accelerati dall'hardware DRAGEN.

Condizioni di funzionamento

Per maggiori informazioni sulle condizioni di funzionamento, consultare la sezione Considerazioni ambientali in *Documentazione del prodotto dello strumento NovaSeq 6000Dx*.

Elemento	Specifica
Temperatura	Mantenere nel laboratorio una temperatura compresa tra 19 °C e 25 °C (22 °C ± 3 °C). Questa temperatura è l'intervallo di 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%. Il sistema deve essere utilizzato a un'altitudine operativa di 2.000 metri o inferiore.

Materiali di consumo e apparecchiature

Questa sezione elenca tutto ciò che è necessario per una corsa di sequenziamento di NovaSeq 6000Dx. Sono compresi i materiali di consumo forniti di Illumina e i materiali di consumo e le apparecchiature accessorie che devono essere acquistati da altri fornitori. Questi elementi sono necessari per completare il protocollo e per eseguire le procedure di manutenzione e risoluzione dei problemi.

Per informazioni sui simboli presenti sui materiali di consumo o sulle confezioni dei materiali di consumo, consultare [Legenda dei simboli IVD Illumina \(documento n. 1000000039141\)](#).

Materiali di consumo per il sequenziamento

Una corsa di NovaSeq 6000Dx richiede i seguenti componenti:

- Cartuccia di tamponi
- Cartuccia con cluster
- Cella a flusso
- Provetta della libreria
- Cartuccia SBS

I materiali di consumo di NovaSeq 6000Dx sono confezionati nelle seguenti configurazioni. Ciascun componente utilizza l'identificazione a radio frequenza (Radio-Frequency IDentification, RFID) per garantire il monitoraggio accurato e la compatibilità dei materiali di consumo.

Tabella 1 Materiali di consumo forniti da Illumina

Nome del kit	Sommario	Numero di catalogo Illumina
Kit reagenti NovaSeq 6000Dx S2 v1.5 (300 cicli)	Cartuccia con cluster S2 Cella a flusso S2 Cartuccia SBS S2	20046931
Kit Reagenti NovaSeq 6000Dx S4 v1.5 (300 cicli)	Cartuccia con cluster S4 Cella a flusso S4 Cartuccia SBS S4	20046933
Cartuccia di tamponi S2 NovaSeq 6000Dx	Cartuccia di tamponi S2	20062292
Cartuccia di tamponi S4 NovaSeq 6000Dx	Cartuccia di tamponi S4	20062293
Provetta della libreria NovaSeq 6000Dx	Provetta della libreria singola	20062290
Provetta della libreria NovaSeq 6000Dx, confezione da 24	24 provette della libreria	20062291

Non appena si ricevono i materiali di consumo, per assicurare un funzionamento corretto provvedere subito a conservare i componenti del kit alla temperatura indicata.

Tabella 2 Conservazione dei kit NovaSeq 6000Dx

Materiale di consumo	Quantità	Temperatura di conservazione	Lunghezza	Larghezza	Altezza
Cella a flusso	1	Tra 2 °C e 8 °C	27,7 cm	17 cm	3,8 cm
Cartuccia con cluster	1	Tra -25 °C e -15 °C	29,5 cm	13 cm	9,4 cm

Materiale di consumo	Quantità	Temperatura di conservazione	Lunghezza	Larghezza	Altezza
Cartuccia SBS	1	Tra -25 °C e -15 °C	30 cm	12,4 cm	11,2 cm
Cartuccia di tamponi	1	Tra 15 °C e 30 °C	42,2 cm	20,6 cm	21,1 cm
Provetta della libreria	1	Tra 15 °C e 30 °C	4,1 cm	2,3 cm	12,4

Dettagli sui materiali di consumo

Per identificare i componenti del kit compatibili, le celle a flusso e le cartucce sono etichettate con simboli che indicano la modalità del kit.

Tabella 3 Etichettatura per la compatibilità

Modalità del kit	Marchio sull'etichetta	Descrizione
Componenti del kit S2		La cella a flusso S2 genera fino a 4,1 miliardi di letture unidirezionali che passano il filtro con un output fino a 1.000 Gb a 2 x 150 bp. La cella a flusso S2 fornisce il sequenziamento rapido per la maggior parte delle applicazioni ad elevata produttività.
Componenti del kit S4		La cella a flusso S4 genera fino a 10 miliardi di letture unidirezionali che passano il filtro con un output fino a 3.000 Gb a 2 x 150 bp. La cella a flusso S4 è una versione a quattro corsie della cella a flusso, progettata per il massimo output.

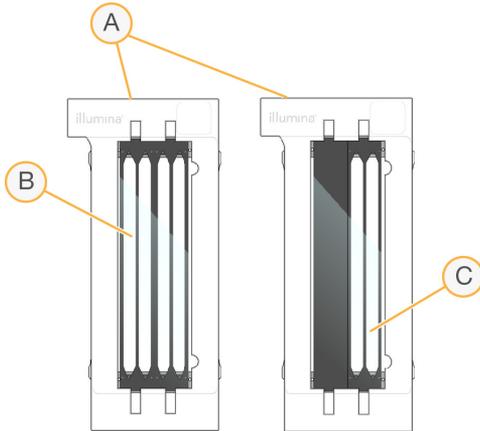
Cella a flusso

La cella a flusso NovaSeq 6000Dx è una cella a flusso preconfigurata inserita in una cartuccia. La cella a flusso è un substrato in vetro che contiene miliardi di nanopozzetti in disposizione ordinata. I cluster sono generati nei nanopozzetti nei quali viene eseguito il sequenziamento.

Ciascuna cella a flusso dispone di più corsie per il sequenziamento di librerie raggruppate in pool. La cella a flusso S2 ha due corsie e la cella a flusso S4 ne ha quattro. Ciascuna corsia viene sottoposta a imaging in strisce multiple, quindi il software divide l'immagine di ciascuna striscia in porzioni più piccole chiamate tile.

Alcuni graffi e altri difetti estetici minori sulla cella a flusso sono normali e si ritiene che non influiscano sulla qualità e sulla resa dei dati. Illumina consiglia di utilizzare queste celle a flusso come di consueto.

Figura 1 Cella a flusso



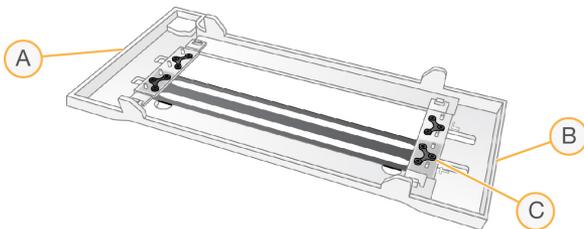
- A. Cartuccia della cella a flusso
- B. Cella a flusso a quattro corsie (S4)
- C. Cella a flusso a due corsie (S2)

La parte inferiore di ciascuna cella a flusso ha guarnizioni multiple. Le librerie e i reagenti entrano nelle corsie della cella a flusso attraverso le guarnizioni sul lato di entrata della cella a flusso. I reagenti usati vengono drenati dalle corsie attraverso le guarnizioni sul lato di uscita.

**ATTENZIONE**

Evitare di toccare le guarnizioni quando si manipola la cella a flusso.

Figura 2 Cella a flusso capovolta



- A. Lato di uscita
- B. Lato di entrata
- C. Guarnizione (una di quattro)

Dettagli delle cartucce di tamponi, con cluster e SBS

Le cartucce di tamponi, con cluster e SBS NovaSeq 6000Dx dispongono di serbatoi sigillati pre-riempiti con reagenti, tamponi e soluzione di lavaggio. Le cartucce Cluster e SBS sono incluse nei kit di reagenti NovaSeq 6000Dx. La cartuccia di tamponi è venduta separatamente.

Le cartucce vengono caricate direttamente sullo strumento, sono codificate a colori ed etichettate per ridurre gli errori di caricamento. Le guide presenti nel cassetto del vano refrigerato per i reagenti e nel cassetto dei tamponi assicurano il corretto orientamento.

Tabella 4 Cartucce NovaSeq 6000Dx

Materiale di consumo	Descrizione
Cartuccia di tamponi 	Pre-riempita con tamponi di sequenziamento e pesa fino a 6,8 kg. Una maniglia in plastica facilita il trasporto, il caricamento e lo scaricamento. La cartuccia di tamponi contiene reagenti sensibili alla luce. Conservare il contenitore dei tamponi imballato fino al momento dell'uso.
Cartuccia con cluster 	Pre-riempita con reagenti PE, di indicizzazione, con cluster e soluzione di lavaggio. Include una posizione designata per la provetta delle librerie. L'etichettatura arancione distingue la cartuccia con cluster dalla cartuccia SBS. Un reagente di denaturazione nella posizione n. 30 contiene formammide, un'amide organica e una tossina riproduttiva. Per semplificare lo smaltimento sicuro di qualsiasi reagente non utilizzato dopo una corsa di sequenziamento, questo serbatoio è rimovibile.
Cartuccia SBS 	Pre-riempita con i reagenti per il sequenziamento a volumi specifici per il numero di cicli supportati dal kit. Ciascuna delle tre posizioni di reagente presenta una posizione riservata adiacente per il lavaggio post-corsa automatico. L'etichettatura grigia distingue la cartuccia SBS dalla cartuccia con cluster. La cartuccia SBS contiene reagenti sensibili alla luce. Conservare il contenitore dell'SBS imballato fino al momento dell'uso.

Serbatoi riservati per cartucce con cluster

Tre serbatoi sono riservati per i primer personalizzati e una posizione vuota è riservata per la provetta delle librerie. Per tracciare i campioni, la provetta delle librerie viene caricata nella cartuccia con cluster durante l'impostazione della corsa e rimane nella cartuccia fino al termine della corsa.

Figura 3 Serbatoi numerati

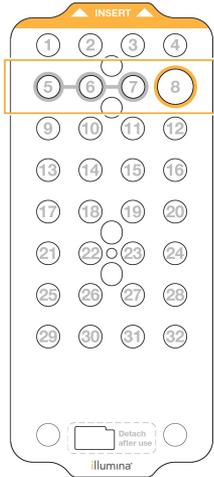


Tabella 5 Serbatoi delle cartucce con cluster

Posizione	Riservata per
5, 6 e 7	Primer personalizzati facoltativi
8	Provetta della libreria

Apparecchiature e materiali di consumo forniti dall'utente

Tabella 6 Materiali di consumo

Materiale di consumo	Fornitore	Scopo
Flacone per centrifuga, 500 ml	Fornitore di laboratorio generico	Diluizione di Tween 20 per un lavaggio di manutenzione.
Provetta per centrifuga, 30 ml	Fornitore di laboratorio generico	Diluizione di NaOCl per un lavaggio di manutenzione.
Guanti monouso, privi di polvere	Fornitore di laboratorio generico	Uso generico.
Salviettine imbevute di alcol isopropilico al 70% oppure Salviettine imbevute di alcol etanolo al 70%	VWR, n. di catalogo 95041-714, o equivalente Fornitore di laboratorio generico	Pulizia dei componenti prima di una corsa e per uso generico.
Panno da laboratorio a bassissimo rilascio di particelle	VWR, n. di catalogo 21905-026, o equivalente	Asciugatura del piano portacelle e per uso generico.

Materiale di consumo	Fornitore	Scopo
NaOCl di grado reagente, 5%	Sigma-Aldrich, n. di catalogo 239305	Esecuzione di un lavaggio di manutenzione.
Punte per pipette da 2 µl	Fornitore di laboratorio generico	Pipettamento per la diluizione e il caricamento delle librerie.
Punte per pipette da 20 µl	Fornitore di laboratorio generico	Pipettamento per la diluizione e il caricamento delle librerie.
Punte per pipette da 200 µl	Fornitore di laboratorio generico	Pipettamento per la diluizione e il caricamento delle librerie.
Punte per pipette da 1.000 µl	Fornitore di laboratorio generico	Pipettamento per la diluizione e il caricamento delle librerie.
Reagente o alcol isopropilico per spettrofotometria (99%), flacone da 100 ml	Fornitore di laboratorio generico	Pulizia periodica dei componenti ottici e supporto della cartuccia di pulizia dell'obiettivo.
Tween 20	Sigma-Aldrich, n. di catalogo P7949	Esecuzione di un lavaggio di manutenzione.
Acqua da laboratorio	Fornitore di laboratorio generico	Diluizione di Tween 20 e di sodio ipoclorito per un lavaggio di manutenzione.

Tabella 7 Apparecchiatura

Apparecchio	Fornitore
Congelatore, tra -25 °C e -15 °C	Fornitore di laboratorio generico
Cilindro graduato, 500 ml, sterile	Fornitore di laboratorio generico
Portaghiaccio	Fornitore di laboratorio generico
Pipette, 20 µl	Fornitore di laboratorio generico
Pipette, 200 µl	Fornitore di laboratorio generico
Pipette, 1000 µl	Fornitore di laboratorio generico
Frigorifero, temperatura compresa tra 2 °C e 8 °C	Fornitore di laboratorio generico
Vasca, bagni d'acqua*	Fornitore di laboratorio generico

* Utilizzare una vasca in grado di contenere due cartucce di reagenti e il corretto livello di acqua. Ad esempio, (61 cm × 91,4 cm × 25,4 cm)

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 Megohm (MΩ)
- Acqua Milli-Q
- Acqua Super-Q
- Acqua sterile per biologia molecolare

Istruzioni per l'uso

Le istruzioni che seguono si riferiscono al funzionamento di Strumento NovaSeq 6000Dx in modalità IVD utilizzando le configurazioni del kit S2 o S4.

Creare una corsa di sequenziamento

Per creare una corsa mediante Illumina Run Manager in modalità IVD o RUO, procedere come segue. In alternativa, selezionare **Import Run** (Importa corsa) nella scheda Planned (Pianificate) della pagina Runs (Corse) e importare un foglio campione. Creare nuove corse sullo strumento o accedendo a Illumina Run Manager con un browser su un computer in rete.

NOTA Le informazioni esatte richieste da ciascuna applicazione di analisi sono diverse, ma il processo di creazione di una corsa comprende le seguenti fasi.

1. Dalla scheda Planned (Pianificate) della schermata Runs (Corse), selezionare **Create Run** (Crea corsa).
2. Selezionare un'applicazione, poi selezionare **Next** (Avanti).
3. Procedere con le schermate di impostazione. A seconda dell'applicazione, le schermate visualizzate possono essere le seguenti:
 - **Run Settings** (Impostazioni della corsa): immettere i parametri della corsa.
 - **Sample Data** (Dati del campione): immettere i dati del campione manualmente o importando un file CSV contenente informazioni sul campione. I nomi dei campioni devono essere univoci.
 - **Analysis settings** (Impostazioni di analisi): immettere le impostazioni per l'analisi.
4. Nella schermata Review (Revisione), rivedere le informazioni sulla corsa e selezionare **Save** (Salva). La corsa viene aggiunta all'inizio dell'elenco delle corse nella scheda Planned (Pianificate).

Preparazione dei materiali di consumo

Scongelamento delle cartucce SBS e con cluster

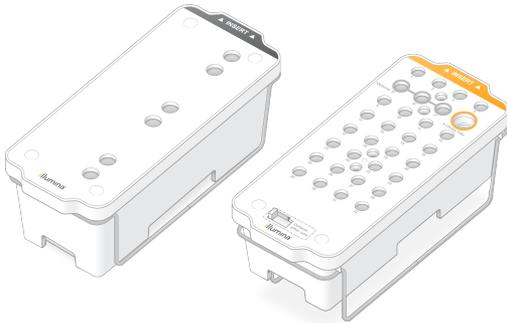


ATTENZIONE

L'utilizzo di acqua calda per lo scongelamento dei reagenti può causare una ridotta qualità dei dati o una corsa non riuscita.

1. Se è in corso una corsa di sequenziamento, assicurarsi che entrambi i lati dello strumento siano disponibili al completamento dello scongelamento.
2. Rimuovere le cartucce SBS e con cluster dalla temperatura di conservazione compresa tra -25°C e -15°C.
3. Posizionare ciascuna cartuccia in una griglia di scongelamento.
I rack sono forniti con lo strumento e impediscono il capovolgimento nel bagno d'acqua.

Figura 4 Cartucce in griglie di scongelamento



4. Fare riferimento alla seguente tabella per determinare la durata dello scongelamento.
Scongelare le cartucce SBS e con cluster a bagnomaria a temperatura ambiente (da 19°C a 25°C) come segue. Immergere le cartucce per circa metà.

Cartuccia	Durata dello scongelamento
Cartuccia SBS S2	4 ore
Cartuccia con cluster S2	Fino a 2 ore
Cartuccia SBS S4	4 ore
Cartuccia con cluster S4	Fino a 4 ore



ATTENZIONE

Il mancato avvio del sequenziamento entro quattro ore dallo scongelamento delle cartucce di reagente può comportare una riduzione della qualità dei dati.

5. Asciugare completamente la base della cartuccia utilizzando fogli di carta. Asciugare tra i pozzetti per rimuovere tutta l'acqua.
6. Ispezionare i sigilli per l'eventuale presenza di acqua. Se è presente acqua, tamponare con un panno che non lascia residui.
7. Ispezionare la parte inferiore di ogni cartuccia per assicurarsi che i recipienti siano privi di ghiaccio, il che significa che i reagenti sono scongelati.
8. Capovolgere ciascuna cartuccia dieci volte per miscelare i reagenti.



ATTENZIONE

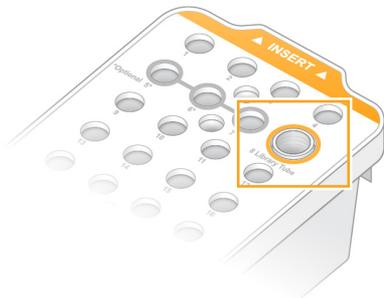
La mancata inversione completa delle cartucce può ridurre la qualità dei dati.

9. Picchiettare delicatamente la parte inferiore di ciascuna cartuccia sul banco per ridurre le bolle d'aria.

Caricamento della provetta delle librerie

1. Senza alterare la libreria nella parte inferiore, inserire la provetta delle librerie senza tappo contenente il pool di librerie denaturato e diluito nella posizione **Library Tube** (Provetta delle librerie) (n. 8) della cartuccia con cluster.
2. Immettere la provetta della libreria nella posizione n. 8 della cartuccia con cluster.

Figura 5 Provetta delle librerie senza tappo caricata nella posizione n. 8



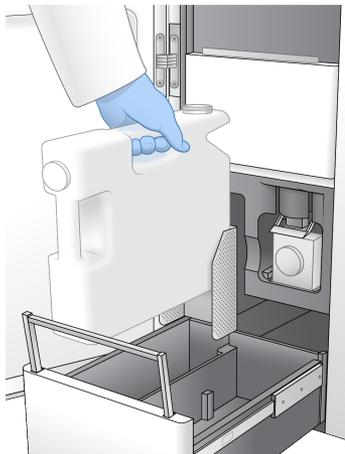
Svuotamento dei flaconi di reagenti usati

Utilizzare le seguenti istruzioni per svuotare i flaconi di reagenti usati con *ogni* corsa di sequenziamento. Se il sistema è configurato per far defluire i reagenti usati all'esterno, il flacone piccolo raccoglie i reagenti usati e deve essere svuotato per ogni corsa di sequenziamento. Il flacone grande deve rimanere in posizione.

1. Rimuovere e svuotare il flacone piccolo dei reagenti usati nel modo seguente.
 - a. Sollevare la leva e rimuovere il flacone piccolo dei reagenti usati dall'alloggiamento. Afferrare il flacone da entrambi i lati.
 - b. Rimuovere il tappo filettato dal supporto dei tappi nella parte anteriore del flacone.

- c. Sigillare l'apertura del flacone con un tappo per impedire le fuoriuscite.
 - d. Mantenendo i contenuti separati dai contenuti di altri flaconi, smaltire in base agli standard applicabili nella propria regione.
 - e. Rimettere il flacone senza tappo nell'alloggiamento, quindi abbassare la leva. Conservare il tappo sul supporto dei tappi.
2. Rimuovere e svuotare il flacone grande dei reagenti usati nel modo seguente.
- a. Utilizzando l'impugnatura superiore, rimuovere il flacone grande dei reagenti usati dal lato sinistro del cassetto dei tamponi.
 - b. Rimuovere il tappo filettato dal supporto dei tappi nella parte anteriore del flacone.
 - c. Sigillare l'apertura del flacone con un tappo per impedire le fuoriuscite.
 - d. Smaltire i contenuti in base agli standard applicabili nella propria regione. Tenere saldamente le impugnature durante lo svuotamento.
 - e. Rimettere il flacone senza tappo nel cassetto dei tamponi. Conservare il tappo sul supporto dei tappi.

Figura 6 Riposizionamento del flacone vuoto nel suo alloggiamento



3. Indossare un nuovo paio di guanti privi di polvere.



ATTENZIONE

Indossare sempre un nuovo paio di guanti dopo aver maneggiato il flacone di reagente usato.

4. Chiudere il cassetto dei tamponi, quindi chiudere gli sportelli dello scomparto dei liquidi.



ATTENZIONE

Se i flaconi di reagenti usati non vengono svuotati si potrebbe verificare una corsa terminata e una fuoriuscita, che potrebbe danneggiare lo strumento e porre un rischio di sicurezza.

Preparazione della cella a flusso

1. Rimuovere dalla confezione una nuova cella a flusso incastolata dalla temperatura di conservazione compresa tra 2°C e 8°C.
2. Tenere da parte la confezione sigillata della cella a flusso a temperatura ambiente (da 19°C a 25°C) per 10-15 minuti.
Utilizzare la cella a flusso entro 12 ore dalla rimozione dalla confezione.

Caricamento dei materiali di consumo

Utilizzare le seguenti istruzioni per avviare l'impostazione della corsa e caricare i materiali di consumo.

1. Dal menu principale selezionare **Sequence** (Sequenziamento) quindi selezionare una corsa con singola cella a flusso o doppia cella a flusso come segue.
 - **A+B**: imposta una corsa con doppia cella a flusso.
 - **A**: imposta una corsa con singola cella a flusso sul lato A.
 - **B**: imposta una corsa con singola cella a flusso sul lato B.Il sistema avvia l'impostazione della corsa, iniziando con il caricamento della cella a flusso.
2. Selezionare **OK** per accettare l'avvertenza e aprire lo sportello della cella a flusso.



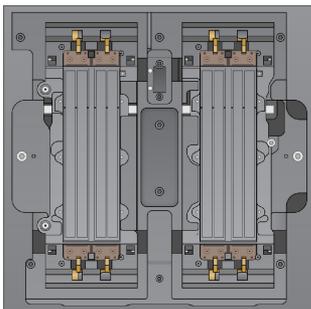
ATTENZIONE

Mantenere la superficie libera durante la corsa di sequenziamento ed evitare di appoggiarsi allo strumento. Premere sullo sportello della cella a flusso per aprirla. Questa azione arresta la corsa. Le corse arrestate non possono essere riprese.

Caricamento della cella a flusso

1. Se presente, rimuovere la cella a flusso utilizzata nella corsa precedente.
2. Se sono visibili particelle sul piano portacelle, pulire l'intero piano, inclusi l'interfaccia della fluidica e la superficie di vetro del target dell'allineamento ottico con una salviettina imbevuta di alcol. Asciugare con un panno che non lascia residui.

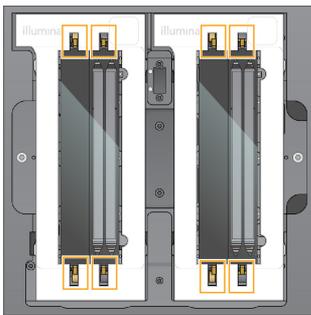
Figura 7 Piano portacelle



3. Rimuovere la cella a flusso dalla confezione come segue.

- a. Indossare un nuovo paio di guanti privi di polvere per evitare di contaminare la superficie in vetro della cella a flusso.
 - b. Tenendo la confezione su una superficie piana, aprirla partendo dal sigillo angolato.
 - c. Rimuovere il contenitore in plastica trasparente che copre la cella a flusso.
 - d. Rimuovere la cella a flusso dalla confezione. Afferrare la cella a flusso dai lati evitando di toccare il vetro o le guarnizioni nella parte inferiore.
 - e. Se sono visibili particelle sulle superfici di vetro, pulire la superficie interessata con una salviettina imbevuta di alcol che non lascia residui e asciugare con un panno da laboratorio a bassissimo rilascio di particelle.
 - f. Eliminare la confezione in base alle procedure di smaltimento.
4. Allineare la cella a flusso sopra i quattro morsetti sollevati e posizionarla sul piano portacelle.

Figura 8 Celle a flusso caricate e allineate sopra i morsetti



5. Selezionare **Close Flow Cell Door** (Chiudi lo sportello della cella a flusso).
Lo sportello della cella a flusso si chiude, i sensori e l'identificazione RFID vengono verificati e l'ID della cella a flusso appare sullo schermo.

Caricamento delle cartucce con cluster e SBS

1. Aprire gli sportelli dello scomparto dei liquidi, quindi aprire lo sportello del vano refrigerato per i reagenti.
2. Rimuovere le cartucce SBS e con cluster usate, se presenti da una corsa precedente.
Le cartucce usate presentano sigilli perforati.
3. Smaltire i contenuti non utilizzati in base agli standard applicabili.
Per il corretto smaltimento della posizione n. 30 della cartuccia con cluster, vedere [Posizione rimovibile n. 30 alla pagina 21](#).

4. Caricare le cartucce preparate nel cassetto del vano refrigerato per i reagenti come segue, in modo che le etichette Insert (Inserisci) siano rivolte verso la parte posteriore dello strumento.
 - Posizionare la cartuccia SBS (etichetta grigia) nella posizione sinistra.
 - Posizionare la cartuccia con cluster (etichetta arancione) contenente la provetta delle librerie senza tappo nella posizione destra.

Figura 9 Cartucce di reagenti caricate



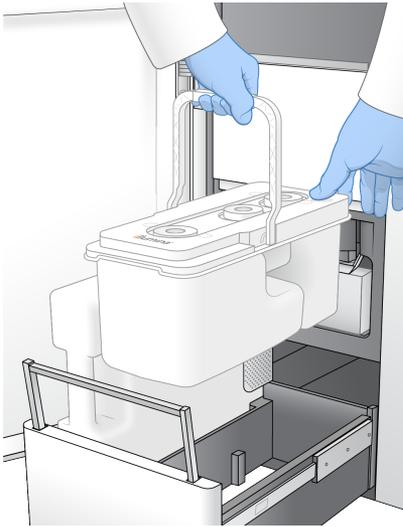
5. Fare scorrere il cassetto nel vano refrigerato per i reagenti, quindi chiudere lo sportello del vano refrigerato. Vengono verificati i sensori e i RFID. Gli ID per la provetta delle librerie e le due cartucce vengono visualizzati sulla schermata.

Caricamento della cartuccia di tamponi

1. Tirare la maniglia in metallo per aprire il cassetto dei tamponi.
2. Rimuovere la cartuccia di tamponi usati che si trova sul lato destro del cassetto dei tamponi. La cartuccia di tamponi usata presenta sigilli perforati.
3. Posizionare una nuova cartuccia di tamponi nel cassetto dei tamponi in modo che l'etichetta Illumina sia rivolta verso la parte anteriore del cassetto. Allineare la cartuccia con le guide sollevate sul piano e sui lati del cassetto.

Quando la cartuccia di tamponi viene caricata e alloggiata correttamente il cassetto si può chiudere.

Figura 10 Caricamento della cartuccia di tamponi



4. Se entrambi i flaconi di reagenti usati sono stati svuotati, selezionare la casella di controllo per confermare che entrambi i flaconi di reagenti usati sono vuoti.

NOTA Se i flaconi di reagenti usati non vengono svuotati si potrebbe verificare una corsa terminata e una fuoriuscita, che potrebbe danneggiare lo strumento e porre un rischio di sicurezza.

5. Una volta aggiunti i materiali di consumo, selezionare **Run Selection** (Esegui selezione) per procedere.

Selezionare e avviare la corsa

Lo strumento esegue una scansione dell'ID della provetta della libreria e cerca una corsa pianificata corrispondente.

1. Se per ogni lato utilizzato viene trovata una corsa pianificata corrispondente all'ID della provetta della libreria, la selezione della corsa viene saltata. Selezionare **Review** (Rivedi) per procedere.
2. Se non ci sono corse corrispondenti per uno o entrambi i lati, selezionare **Run Selection** (Selezione corsa), quindi selezionare una o più corse pianificate.
Non è possibile selezionare la stessa corsa pianificata su entrambi i lati.
3. Quando sono selezionate una o più corse, selezionare **Pre-Run Checks** (Controlli pre-corsa).
4. Il completamento della verifica pre-corsa impiega circa cinque minuti.
Dopo il corretto completamento, la corsa viene avviata automaticamente.

NOTA Per evitare l'eccessivo caricamento del disco rigido, non copiare alcun dato su C:\ dopo l'avviamento della corsa.

Errori della verifica pre-corsa

1. Se le verifiche pre-corsa non viene superata a causa di un errore dei sensori, come una cella a flusso non rilevata, è necessario uscire e riavviare il flusso di lavoro.
2. Se altre verifiche pre-corsa non vengono superate, selezionare **Retry** (Riprova) per riavviare le verifiche non superate o **Retry All** (Riprova tutte) per riavviare tutte le verifiche.
Gli errori devono essere risolti prima di poter avviare la corsa.
3. Selezionare l'icona **Errore** per visualizzare i dettagli relativi all'errore.
4. Se la verifica dell'allineamento non viene superata, risolvere l'errore nel modo seguente.
 - a. Selezionare **Reload** (Ricarica), quindi selezionare **OK** per confermare e tornare alla schermata Load (Carica).
 - b. Rimuovere qualsiasi oggetto che si trovi nella parte superiore dello strumento, quindi selezionare **OK**. Lo sportello della cella a flusso si apre.
 - c. Ricaricare la cella a flusso, quindi selezionare **Run Setup** (Impostazione corsa).
 - d. Completare ciascuna schermata per rileggere ogni RFID e tornare alla schermata Pre-Run Checks (Verifiche pre-corsa).
 - e. Rieseguire la verifica.

Monitoraggio del progresso della corsa

Nella schermata Sequencing (Sequenziamento) vengono visualizzati i seguenti dettagli mentre la corsa è in atto. La schermata Sequencing (Sequenziamento) è accessibile dal menu principale.

- **Stato delle singole fasi della corsa**
- **Time to completion** (Data e ora del completamento): la data e l'ora in cui è stata completata la corsa (aaaa-mm-gg hh:mm).
- **Run progress** (Progresso della corsa): l'attuale fase della corsa. La dimensione della barra di progresso non è proporzionale alla velocità della corsa di ciascuna fase.
- **Q-scores** (Punteggi qualitativi): la distribuzione dei punteggi qualitativi.
- **Intensity** (Intensità): il valore delle intensità dei cluster per il 90° percentile per ciascuna tile. I colori del grafico indicano i canali rosso e verde.
- **Clusters Passing Filter (%)** (Cluster che attraversano il filtro - %): la percentuale di cluster che attraversano il filtro.
- **Projected Total Yield (GB)** (Resa totale prevista (GB)): la resa prevista per la corsa della cella a flusso. Se sono state selezionate le metriche (H) i numeri visualizzati corrispondono alla resa attuale per corsia e all'aggiornamento per ogni ciclo per tutta la durata della corsa.
- **Q30** (Punteggio qualitativo 30): la percentuale di identificazioni delle basi per la corsa che ottengono un punteggio qualitativo ≥ 30 .

Icone di stato

Lo stato della corsa è indicato da un'icona di stato sull'interfaccia NVOS. Un numero sull'icona indica il numero di condizioni per uno stato.

Quando uno stato della corsa viene modificato, l'icona lampeggia. Selezionare l'icona per visualizzare una descrizione della condizione. Selezionare **Acknowledge** (Accetta) per cancellare il messaggio, quindi **Close** (Chiudi) per chiudere la finestra di dialogo.

Icona di stato	Nome dello stato	Descrizione
	Stato OK	Le condizioni del sistema sono normali.
	Elaborazione	Il sistema è in fase di elaborazione.
	Avvertenza	Si è verificata un'avvertenza che richiede attenzione. Le avvertenze non arrestano una corsa o richiedono un intervento prima di poter procedere.
	Errore	Si è verificato un errore. Gli errori richiedono un intervento prima di poter procedere con la corsa.
	Informazione	È disponibile un messaggio non urgente.

Metriche della corsa

Il software visualizza le metriche generate durante la corsa. Le metriche vengono visualizzate sotto forma di grafici, diagrammi e tabelle in base ai dati generati da RTA3 e scritti nei file InterOp.

La clusterizzazione dura circa 2 ore, quindi il sequenziamento inizia con il ciclo 1. Le metriche vengono aggiornate man mano che il sequenziamento procede. I cluster che attraversano il filtro, la resa e i punteggi qualitativi sono disponibili dopo il ciclo 26. Prima del ciclo 26, nessun valore viene popolato e vengono indicati come non applicabile.

Dopo il sequenziamento

Le sezioni seguenti forniscono istruzioni sulle fasi successive al completamento del sequenziamento.

Lavaggio post-corsa automatico

Quando viene completato il sequenziamento, il software avvia un lavaggio post-corsa automatico che richiede circa 80 minuti. Il sistema pompa ipoclorito sodio (NaOCl) allo 0,24% dalla posizione n. 17 e lo diluisce allo 0,12%. NaOCl allo 0,12% viene pompato nel reagente ExAmp e nelle posizioni della libreria, attraverso la cella a flusso, quindi ai flaconi di reagente usato. Il lavaggio risciacqua il modello dal sistema per impedire la contaminazione incrociata.

Al termine del lavaggio, il sistema viene messo in uno stato di sicurezza e il pulsante Home (Inizio) diventa attivo. Lasciare i materiali di consumo in posizione fino alla corsa successiva. Dopo il lavaggio, i pescanti rimangono nelle cartucce SBS e con cluster per impedire che l'aria entri nel sistema. I pescanti nella cartuccia di tamponi vengono sollevati in modo che i flaconi dei reagenti usati possano essere svuotati. Il tampone di lavaggio viene quindi pompato attraverso tutte le linee per eliminare il NaOCl e i reagenti dal sistema.

NOTA Se si verifica un errore durante un lavaggio post-corsa automatico, e il lavaggio post-corsa è incompleto, è richiesto un lavaggio di manutenzione.

Posizione rimovibile n. 30

Il serbatoio in posizione n. 30 della cartuccia con cluster contiene formammide. Il serbatoio viene rimosso dalla cartuccia con cluster usata e smaltito separatamente.



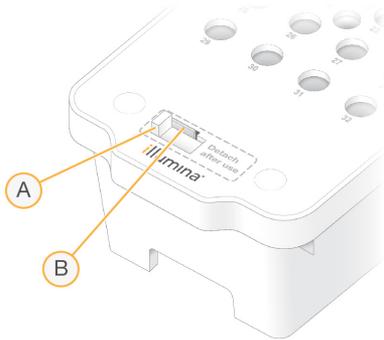
ATTENZIONE

Questo set di reagenti contiene materiali 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 pertinenti a livello regionale, nazionale e locale. Per ulteriori informazioni ambientali, di salute e di sicurezza, consultare le SDS alla pagina Web support.illumina.com/sds.html.

1. Indossando un paio di guanti, spingere la linguetta in plastica bianca etichettata **Detach after use** (Rimuovere dopo l'uso) che si trova sulla destra.
2. Posizionare una mano o una superficie sotto il serbatoio, premere la linguetta in plastica trasparente verso l'etichetta Illumina per rilasciare il serbatoio da sotto la cartuccia con cluster.

NOTA Evitare di sovrapporre le cartucce con cluster durante la conservazione. La sovrapposizione potrebbe causare il distacco accidentale del serbatoio.

Figura 11 Posizione rimovibile n. 30



- A. Linguetta in plastica bianca per la rimozione
- B. Linguetta in plastica trasparente per il rilascio

3. Smaltire il serbatoio in base agli standard applicabili.

Output del sequenziamento

Durante il sequenziamento, i dati vengono trasferiti automaticamente da Strumento NovaSeq 6000Dx a Illumina DRAGEN Server. Quando l'analisi primaria è terminata e il trasferimento dei dati è completo, l'analisi secondaria su Illumina DRAGEN Server può iniziare automaticamente utilizzando le opzioni di analisi definite dall'applicazione selezionata in Illumina Run Manager. I risultati prodotti dipendono dalle opzioni scelte durante l'impostazione della corsa. Per visualizzare i risultati di una corsa, selezionare il nome della corsa desiderata nella scheda Completed (Completata) della schermata Runs (Corse). I file di output si trovano anche nella posizione specificata nella schermata Instrument Settings (Impostazioni strumento).

Analisi in tempo reale

Strumento NovaSeq 6000Dx utilizza RTA3, un'implementazione del software Analisi in tempo reale, sul Compute Engine (CE) dello strumento. RTA3 estrae le intensità dalle immagini ricevute dalla fotocamera, esegue l'identificazione delle basi, assegna un punteggio di qualità all'identificazione delle basi, si allinea a PhiX e riporta i dati in file InterOp.

Per ottimizzare il tempo di elaborazione, RTA3 archivia le informazioni in memoria. Se RTA3 viene terminato, l'elaborazione non riprende e tutti i dati della corsa elaborata archiviati in memoria vengono persi.

Input RTA3

RTA3 richiede immagini di tile contenute nella memoria locale del sistema per l'elaborazione. RTA3 riceve informazioni e comandi di esecuzione da NVOS.

Output RTA3

Le immagini per ciascun canale colore sono passate in memoria a RTA3 come tile. In base a queste immagini, RTA3 produce output sotto forma di un set di file di identificazione delle basi qualitativamente valutate e di file filtro. Tutti gli altri output sono file di output di supporto.

Tipo di file	Descrizione
File di identificazione delle basi	Ciascuna tile analizzata viene inclusa in un file di identificazione delle basi concatenato (*.cbcl). Le tile appartenenti alla stessa corsia e superficie sono aggregate in un CBCL per ciascuna corsia e superficie.
File filtro	Ciascuna tile produce un file filtro (*.filter) che specifica se un cluster ha attraversato il filtro.

RTA3 fornisce metriche in tempo reale della qualità della corsa sotto forma di file InterOp, ossia output binari che contengono metriche su tile, ciclo e a livello di lettura.

Gestione degli errori

RTA3 crea file di registro e li scrive nella cartella Logs (Registri). Gli errori vengono registrati in un file di testo in formato file *.log.

I seguenti file di registro sono trasferiti alla destinazione di output finale al termine dell'elaborazione:

- `info_00000.log` riepiloga importanti eventi di una corsa.
- `error_00000.log` elenca gli errori che si sono verificati durante una corsa.
- `warning_00000.log` elenca le avvertenze che si sono verificate durante una corsa.

Tile della cella a flusso

Le tile sono piccole aree di imaging sulla cella a flusso. La videocamera acquisisce una singola immagine di ciascuna striscia, che il software divide in tile per l'elaborazione di RTA3. Il numero totale di tile dipende da quante corsie, strisce e superfici sono state sottoposte a imaging sulla cella a flusso.

- Le celle a flusso S2 dispongono di 1.408 tile.
- Le celle a flusso S4 dispongono di 3.744 tile.

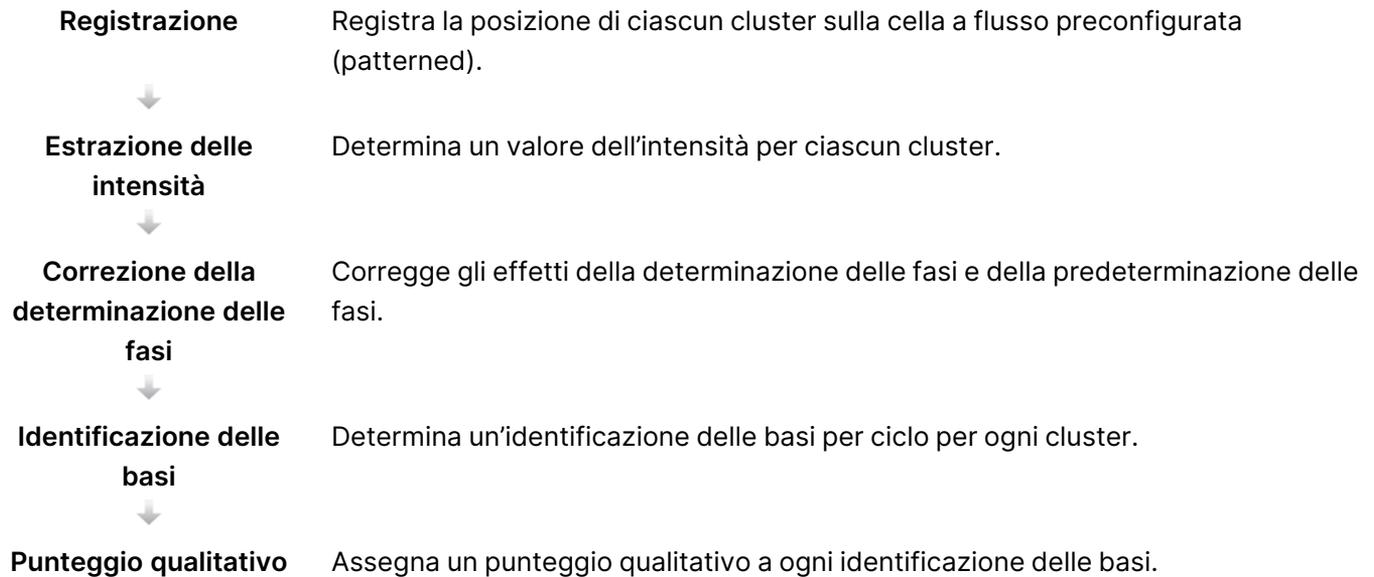
Componente della cella a flusso	S2	S4	Descrizione
Corsie	2	4	Una corsia è un canale fisico con porte di ingresso e di uscita.

Componente della cella a flusso	S2	S4	Descrizione
Superfici	2	2	Le celle a flusso S2 e S4 sono sottoposte a imaging su due superfici: la superficie superiore e la superficie inferiore. La superficie superiore di una tile viene sottoposta a imaging per prima.
Strisce per corsia	4	6	Una striscia è una colonna in una corsia della cella a flusso che la videocamera cattura come un'immagine scansionata.
Tile per striscia	88	78	Una tile è una porzione di una striscia e rappresenta un'area sottoposta a imaging sulla cella a flusso.
Tile totali generate	1.408	3.744	Corsie × superfici × strisce × tile per striscia equivalgono al numero totale di tile.

Il nome della tile è un numero di cinque cifre che rappresenta la posizione sulla cella a flusso. Ad esempio, il nome della tile 1_1205 indica corsia numero 1, superficie superiore, striscia numero 2, tile numero 5.

- La prima cifra rappresenta il numero della corsia:
 - 1 o 2 per una cella a flusso S2.
 - 1, 2, 3 o 4 per una cella a flusso S4.
- La seconda cifra rappresenta la superficie: 1 per superficie superiore e 2 per superficie inferiore.
- La terza cifra rappresenta il numero di striscia:
 - 1, 2, 3 o 4 per una cella a flusso S2.
 - 1, 2, 3, 4, 5 o 6 per una cella a flusso S4.
- Le ultime due cifre rappresentano il numero della tile. La numerazione inizia da 01 sul lato di uscita della cella a flusso, proseguendo fino a 88 o 78 sul lato di presa.
 - Da 01 a 88 per una cella a flusso S2.
 - Da 01 a 78 per una cella a flusso S4.

Flusso di lavoro di Real-Time Analysis



Registrazione

La registrazione allinea un'immagine su un array quadrato ruotato di nanopozzetti su una cella a flusso preconfigurata (patterned). Grazie alle posizioni ordinate dei nanopozzetti, le coordinate X e Y per ogni cluster in una tile sono predeterminate. Le posizioni dei cluster sono scritte su un file di posizioni cluster (s.locs) per ogni corsa.

Se la registrazione non riesce per una qualsiasi immagine in un ciclo, non viene generata alcuna identificazione delle basi per quella tile in quel ciclo.

Estrazione delle intensità

Dopo la registrazione, l'estrazione delle intensità calcola il valore dell'intensità per ogni nanopozzetto in una data immagine. Se la registrazione non riesce, l'intensità per quella tile non può essere estratta.

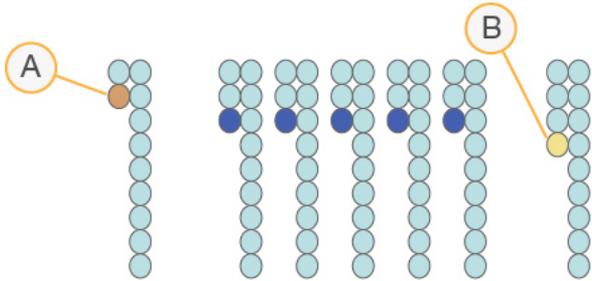
Correzione della determinazione delle fasi

Durante la reazione di sequenziamento, ciascun filamento di DNA in un cluster si estende di una base per ciclo. La determinazione delle fasi e la predeterminazione delle fasi si verificano quando un filamento fuoriesce dalla fase con il ciclo di incorporazione attuale.

La determinazione delle fasi si verifica quando l'incorporazione di una base rimane indietro.

La predeterminazione delle fasi si verifica quando l'incorporazione una base salta in avanti.

Figura 12 Determinazione delle fasi e predeterminazione delle fasi



- A. Lettura con una base nella determinazione delle fasi
- B. Lettura con una base nella predeterminazione delle fasi

RTA3 corregge gli effetti della determinazione delle fasi e della predeterminazione delle fasi che massimizza la qualità dei dati a ogni ciclo per tutta la corsa.

Identificazione delle basi

L'identificazione delle basi determina una base (A, C, G o T) per ogni cluster di una data tile a un ciclo specifico. Strumento NovaSeq 6000Dx utilizza il sequenziamento a due canali, che richiede solo due immagini per codificare i dati per quattro basi di DNA, un'immagine dal canale verde e una dal canale rosso.

Una mancata identificazione viene indicata con una N. Le mancate identificazioni si verificano quando un cluster non attraversa il filtro, non viene eseguita la registrazione o un cluster si è spostato al di fuori dell'immagine.

Le intensità di ciascun cluster sono estratte dalle immagini del canale rosso e del canale verde e sono confrontate tra di loro fornendo quattro popolazioni distinte. Ogni popolazione corrisponde a una base. Il processo di identificazione delle basi determina a quale popolazione appartiene ciascun cluster.

Figura 13 Visualizzazione delle intensità dei cluster

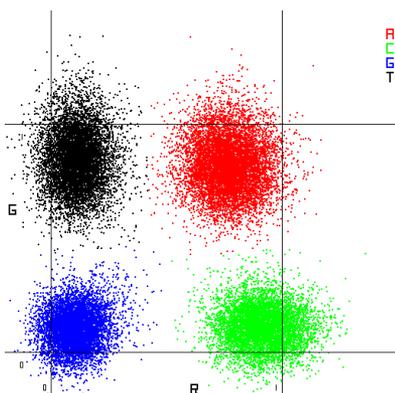


Tabella 8 Identificazione delle basi nel sequenziamento a due canali

Base	Canale rosso	Canale verde	Risultato
A	1 (on)	1 (on)	Cluster che mostrano intensità sia nel canale rosso che nel canale verde.
C	1 (on)	0 (off)	Cluster che mostrano intensità solo nel canale rosso.
G	0 (off)	0 (off)	I cluster che non mostrano intensità a una posizione cluster nota.
T	0 (off)	1 (on)	Cluster che mostrano intensità solo nel canale verde.

Cluster che attraversano il filtro

Durante la corsa, RTA3 filtra i dati non elaborati e rimuove le letture che non soddisfano la soglia per la qualità dei dati. I cluster sovrapposti o di bassa qualità vengono rimossi.

Per l'analisi a due canali, RTA3 utilizza un sistema basato sulla popolazione per determinare il valore chastity (misurazione della purezza dell'intensità) di un'identificazione delle basi. I cluster attraversano il filtro (PF) quando non più di un'identificazione delle basi nei primi 25 cicli presenta un valore chastity inferiore alla soglia fissata. Quando incluso, l'allineamento PhiX viene eseguito al ciclo 26 su un sottogruppo di tile per i cluster che hanno attraversato il filtro. I cluster che non attraversano il filtro non sono identificati come basi e non vengono allineati.

Punteggi qualitativi

Un punteggio qualitativo (Q-score) è una previsione della probabilità di un'identificazione delle basi errata. Un punteggio qualitativo superiore implica che un'identificazione delle basi presenta una qualità superiore ed è più probabile che sia corretta. Dopo la determinazione del punteggio qualitativo, i risultati sono registrati nei file CBCL.

Il punteggio qualitativo fornisce in modo succinto le probabilità di piccoli errori. I punteggi qualitativi sono rappresentati come Q(X), dove X è il punteggio. La tabella seguente illustra la relazione fra un punteggio qualitativo e la probabilità di errore.

Punteggio qualitativo Q(X)	Probabilità di errore
Q40	0,0001 (1 su 10.000)
Q30	0,001 (1 su 1.000)
Q20	0,01 (1 su 100)
Q10	0,1 (1 su 10)

Punteggio qualitativo e report

Il punteggio qualitativo calcola un set valori per ciascuna identificazione delle basi, quindi utilizza questi valori per individuare il punteggio qualitativo in una tabella qualitativa. Le tabelle qualitative sono create per fornire previsioni di qualità accurate e ottimali per le corse generate da una specifica configurazione di una piattaforma di sequenziamento e versione della chimica.

Il punteggio qualitativo si basa su una versione modificata dell'algoritmo Phred.

Per generare la tabella Q per Strumento NovaSeq 6000Dx, vengono identificati tre gruppi di identificazioni delle basi, in base ai cluster di queste specifiche caratteristiche predittive. Dopo il raggruppamento delle identificazioni delle basi, la percentuale di errore media è stata calcolata empiricamente per ognuno di questi tre gruppi e i corrispondenti punteggi qualitativi sono stati registrati nella tabella Q assieme alle caratteristiche predittive relative a quel gruppo. In questo modo, con RTA3 sono possibili solo tre punteggi qualitativi e questi punteggi qualitativi rappresentano la percentuale di errore media del gruppo. Nel complesso, questo risultato è semplificato, sebbene il punteggio qualitativo sia altamente accurato. I tre gruppi nella tabella di qualità corrispondono alle identificazioni delle basi marginali (inferiore a Q15), medie (circa Q20) e di elevata qualità (superiore a Q30) e vengono assegnati punteggi specifici di 12, 26 e 34 rispettivamente. Inoltre, un punteggio nullo di 2 viene assegnato a ogni identificazione non riuscita. Questo modello per riportare i punteggi qualitativi riduce lo spazio di archiviazione e i requisiti di ampiezza di banda senza incidere sull'accuratezza o sulle prestazioni.

Figura 14 Punteggio qualitativo semplificato con RTA3



File di output per il sequenziamento

Tipo di file	Descrizione, posizione e nome del file
File di identificazione delle basi	Ciascun cluster analizzato viene incluso in un file delle identificazione delle basi, aggregato in un file per ciclo, corsia e superficie. Il file aggregato contiene l'identificazione delle basi e il punteggio qualitativo codificato per ogni cluster. Data\Intensities\BaseCalls\L001\C1.1 L[lane]_[surface].cbcl, per esempio L001_1.cbcl
File posizione cluster	Per ciascuna cella a flusso, un file binario della posizione dei cluster contiene le coordinate XY per i cluster in una tile. Un layout esagonale che corrisponde al layout dei nanopozzetti della cella a flusso predefinisce le coordinate. Data\Intensities s_[lane].locs
File filtro	I file filtro specificano se un cluster ha attraversato i filtri. I file filtro sono generati al ciclo 26 utilizzando 25 cicli di dati. Per ogni tile, viene generato un file filtro. Data\Intensities\BaseCalls\L001 s_[lane]_[tile].filter
File informazioni corsa	Elenca il nome della corsa, il numero di cicli in ciascuna lettura, se la lettura è un Index Read (Lettura Indici) e il numero di strisce e tile sulla cella a flusso. Il file informazioni corsa viene creato all'inizio della corsa. [Root folder],RunInfo.xml
File immagini in miniatura	Immagini in miniatura per il primo ciclo di ciascuna lettura di sequenziamento. Thumbnail_Images\L001\C[X.1]: i file sono conservati in una sottocartella per ciascun ciclo. s_[lane]_[tile]_[channel].jpg: l'immagine in miniatura include il numero della tile.

Struttura della cartella di output del sequenziamento

NVOS genera automaticamente il nome della cartella di output.

 **Config** (Configurazione): le impostazioni di configurazione per la corsa.

 **Logs** (Registri): i file di registro che descrivono le fasi operative, l'analisi dello strumento e gli eventi di RTA3.

 SampleSheet.csv: il foglio campioni o altri file allegati, se applicabile.

 **Data** (Dati)

 **Intensities** (Intensities)

 **BaseCalls**

 **L00[X]**: i file di identificazione delle basi (*.cbcl) aggregati in un file per corsia, superficie e ciclo.

 s.locs: il file delle posizioni dei cluster per la corsa.

- 📁 **InterOp**: file binari.
- 📁 **Recipe** (Ricetta): il file della ricetta specifico per la corsa.
- 📁 **Thumbnail Images** (Miniature): le immagini in miniatura ogni decima tile.
- 📁 **LIMS**: il file di impostazione della corsa (*.json), se applicabile.
- 📁 **Audit** (Ispezione)
 - 📄 AuditInfo.xml
- 📄 RTA3.cfg
- 📄 RunInfo.xml
- 📄 RunParameters.xml
- 📄 RTAComplete.txt
- 📄 CopyComplete.txt
- 📄 SequenceComplete.txt
- 📄 IlluminaRunManagerCopyComplete.txt
- 📄 Manifest.tsv

Avvertenze e precauzioni



ATTENZIONE

La legge federale limita la vendita di questo dispositivo da parte o dietro prescrizione di un medico o di un medico autorizzato dalla legge dello stato in cui esercita, ad usare o ad ordinare l'uso del dispositivo.

- **Alcuni componenti dei reagenti forniti da Illumina da utilizzare con Strumento NovaSeq 6000Dx 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 pertinenti a livello regionale, nazionale e locale.** Per ulteriori informazioni sulle considerazioni ambientali, sulla sicurezza e sulla salute, fare riferimento alle schede di sicurezza (Safety Data Sheet, SDS) alla pagina Web support.illumina.com/sds.html.
- In caso contrario le procedure indicate potrebbero fornire risultati errati o una significativa riduzione nella qualità del campione.
- 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.

- 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.
- 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, blocchi termici, agitatori e centrifughe).
- L'indice per l'associazione dei campioni deve corrispondere esattamente all'indice del layout della piastra. L'applicazione DNA Prep with Enrichment compila automaticamente gli index primer associati ai nomi dei campioni, quando vengono inseriti durante la configurazione della corsa. 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.
- Installazione Si consiglia vivamente un software antivirus, fornito dall'utente, per proteggere il computer dai virus.
- Non utilizzare NovaSeq 6000Dx 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.
- 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.
- Lo strumento pesa circa 480 kg ed eventuali cadute o errata manipolazione possono causare lesioni gravi.

Caratteristiche delle prestazioni

Le caratteristiche prestazionali per lo strumento NovaSeq 6000Dx sono state stabilite utilizzando DNA Prep with Enrichment Dx di Illumina per la preparazione delle librerie, Kit reagenti NovaSeq 6000Dx S2 v1.5 (300 cicli) e Kit Reagenti NovaSeq 6000Dx S4 v1.5 (300 cicli) per il sequenziamento e l'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina per l'analisi secondaria, compreso il rilevamento delle varianti della linea germinale e somatiche. 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. Per le caratteristiche prestazionali relative ai fattori pre-analitici, come i metodi di estrazione o le sostanze interferenti, vedere *Inserto della confezione di DNA Prep with Enrichment Dx di Illumina*.

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.

- $(\text{n. di loci delle varianti riportato correttamente dal saggio}) / (\text{n. totale di loci delle varianti})$

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 test.

- $(\text{n. di loci wild type riportato correttamente dal test}) / (\text{n. totale di loci wild type})$

I loci wild type riportati dal test 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.

- $[(\text{n. di loci delle varianti riportato correttamente dal test}) + (\text{n. di loci wild type riportato correttamente dal test})] / [(\text{n. totale di loci delle varianti}) + (\text{n. totale di loci wild type})]$

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 identificazioni positive (Percent Positive Calls, PPC) è il numero di osservazioni con la variante rilevata diviso per il numero totale di osservazioni analizzate, escludendo le osservazioni non valide o quelle filtrate come a bassa profondità.
6. La percentuale di identificazioni negative (Percent Negative Calls, PNC) è calcolata come il numero di osservazioni con rapido riferimento come esito in una posizione diviso per il numero totale di osservazioni testate escludendo le osservazioni non valide o quelle filtrate come a bassa profondità.
7. La percentuale di calcolabilità degli autosomi è calcolata come la percentuale di posizioni di riferimento non-N in regioni mirate in cromosomi autosomici con un'identificazione genotipica positiva.

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 (12) e massimo (192) che Strumento NovaSeq 6000Dx è in grado di sequenziare in una singola corsa di sequenziamento. Sono stati analizzati dodici campioni univoci Platinum Genome DNA (NA12877–NA12888) con almeno 12 diverse combinazioni di index primer per campione. Le librerie di campioni sono state preparate utilizzando un test rappresentativo progettato per interrogare una serie di geni che coprono 1.970.505 basi su tutti i 23 cromosomi umani. I risultati dei campioni di quattro sequenziamenti effettuati con il flusso di lavoro di analisi Germline della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina sono stati confrontati con la versione 2016-1.0 di Platinum Genome.

Per la prima serie di corse, 192 librerie di campioni indicizzati in modo univoco sono state sequenziate in due corse di sequenziamento, una ciascuna con i reagenti S2 e S4, per verificare sia il numero massimo di indici supportati sia la capacità del test di effettuare in modo coerente un'identificazione di genotipizzazione per un dato campione attraverso diverse combinazioni di primer di indicizzazione. Per il secondo set di corse, sono state analizzate 12 librerie di campioni indicizzati univocamente in due corse di sequenziamento, ciascuna con i reagenti S2 e S4, al fine di verificare il numero minimo di indici supportati.

Per le corse a 192 indici, la PPA per SNV variava dal 99,7% al 100%, la PPA per le inserzioni era del 100%, la PPA per le delezioni variava dal 96,7% al 100% e l'NPA era del 100%. Per le corse a 12 indici, la PPA per le SNV variava dal 99,7% al 100%, la PPA per le inserzioni dall'89,6% al 100%, la PPA per le delezioni dal 94,6% al 100% e l'NPA era del 100%.

Carry-over dei campioni

Strumento NovaSeq 6000Dx 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). Dodici campioni di DNA Platinum Genome, sei maschili e sei femminili, sono stati analizzati con un test rappresentativo disegnato per interrogare una serie di geni che coprono 1.970.505 basi su tutti i 23 cromosomi umani, compresi entrambi i cromosomi sessuali. Le librerie sono state sequenziate su Strumento NovaSeq 6000Dx utilizzando flusso di lavoro di analisi Germline della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina. È stato osservato il carry-over dei campioni maschili nei campioni femminili in presenza di letture del target 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 l'analisi del carry-over del campione entro una corsa di sequenziamento, un pool di librerie costituito da almeno dodici repliche di ciascun campione univoco maschile e femminile più due controlli senza template, per un totale di 192 librerie indicizzate in modo univoco, è stato sequenziato su Strumento NovaSeq 6000Dx in due corse di sequenziamento, ciascuna con i reagenti S2 e S4. Il carry-over dei campioni entro una corsa è stato valutato confrontando la copertura del target del cromosoma Y di ciascun replicato femminile sulla copertura media del target del cromosoma Y di tutti i replicati maschili presenti nel raggruppamento in pool. Il 95° percentile del carry-over entro la corsa osservato è stato dello 0,0090% e dello 0,041% per i reagenti S2 e S4, rispettivamente.

Per l'analisi del carry-over dei campioni da una corsa all'altra, sono stati preparati due pool di librerie e sequenziati consecutivamente su un Strumento NovaSeq 6000Dx, con il lato A che utilizzava i reagenti S4 e il lato B che utilizzava i reagenti S2. Il primo pool conteneva almeno dodici repliche di sei campioni femminili univoci più due controlli senza modello, per un totale di 96 librerie indicizzate in modo univoco. Il secondo pool conteneva almeno dodici repliche di sei campioni maschili univoci più due controlli senza modello, per un totale di 96 librerie indicizzate in modo univoco. 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 per tipo di reagente, S2 e S4, confrontando la copertura del target

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. Il 95° percentile del carry-over osservato da una corsa all'altra è stato dello 0,0089% e dello 0,012% per i reagenti S2 e S4, rispettivamente.

Input di DNA

Sangue (Germline)

L'intervallo di input del DNA ematico per il kit DNA Prep with Enrichment Dx di Illumina utilizzando l'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina è stato stabilito per NovaSeq 6000Dx. Questo è stato valutato eseguendo uno studio di diluizione seriale utilizzando otto campioni di DNA Platinum Genome (NA12877 - NA12884) con un test rappresentativo disegnato per interrogare una serie di geni che coprono 1.970.505 basi su tutti i 23 cromosomi umani. Le librerie sono state sequenziate su un Strumento NovaSeq 6000Dx usando un lotto ciascuno di Kit reagenti NovaSeq 6000Dx S2 v1.5 (300 cicli) e Kit Reagenti NovaSeq 6000Dx S4 v1.5 (300 cicli).

Sette campioni sono stati analizzati in duplicati a sei livelli di input di DNA che andavano da 1.000 ng a 10 ng (1.000 ng, 250 ng, 100 ng, 50 ng, 25 ng e 10 ng). Un ottavo campione (NA12884) è stato analizzato come singolo replicato a 10 ng di input e in duplicato per tutti gli altri livelli di input. Per la determinazione dell'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 presentata in [Risultati della PPA per ciascun input di DNA ematico per tipo di variante alla pagina 34](#). La NPA è presentata in [NPA per ciascun input di DNA ematico alla pagina 35](#). Tutti i livelli di input presentano un'accuratezza simile. L'input di DNA ematico raccomandato per DNA Prep with Enrichment Dx di Illumina è di 50-1.000 ng, dove 1.000 ng e 10 ng forniscono un limite superiore e inferiore per soddisfare le caratteristiche prestazionali quando sequenziato su NovaSeq 6000Dx.

Tabella 9 Risultati della PPA per ciascun input di DNA ematico per tipo di variante

Input di DNA (ng)	Tipo di variante	Varianti previste	TP	FN	Identificazioni varianti non rilevate	PPA (%)
10	SNV	69612	69538	68	6	99,9
25		74192	74105	75	12	99,9
50		74105	74	13	99,9	
100		74116	72	4	99,9	
250		74113	72	7	99,9	
1000		74112	73	7	99,9	

Input di DNA (ng)	Tipo di variante	Varianti previste	TP	FN	Identificazioni varianti non rilevate	PPA (%)
10	Inserzione	2732	2732	0	0	100
25		2928	2916	6	6	99,8
50			2914	8	6	99,7
100			2917	6	5	99,8
250			2928	0	0	100
1000			2921	5	2	99,8
10	Delezione	2084	2049	4	31	99,8
25		2240	2200	9	31	99,6
50			2207	3	30	99,9
100			2199	1	40	>99,9
250			2201	0	39	100
1000			2195	2	43	99,9

Tabella 10 NPA per ciascun input di DNA ematico

Input di DNA (ng)	TN	FP	Identificazioni riferimenti non rilevate	NPA (%)
10	115449045	384	285751	>99,9
25	123012157	415	438153	>99,9
50	122985299	369	465043	>99,9
100	122976660	321	473730	>99,9
250	122971099	331	479289	>99,9
1000	122978527	324	471882	>99,9

FFPE (Somatic)

L'intervallo di input del DNA fissato in formalina e incluso in paraffina (Formalin-Fixed Paraffin-Embedded, FFPE) per il kit DNA Prep with Enrichment Dx di Illumina utilizzando l'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina è stato stabilito per NovaSeq 6000Dx. Questo è stato valutato eseguendo uno studio di diluizione in serie utilizzando due campioni Platinum Genome con un test rappresentativo disegnato per investigare una varietà geni che coprono 1.970.505 basi su 23 cromosomi umani. Le librerie sono state sequenziate su un Strumento NovaSeq 6000Dx usando un lotto ciascuno di Kit reagenti NovaSeq 6000Dx S2 v1.5 (300 cicli) e Kit Reagenti NovaSeq 6000Dx S4 v1.5 (300 cicli).

Il campione di DNA GM12877 è stato diluito con il campione di DNA GM12878 per creare GM12877-13 con varianti eterozigoti e omozigoti univoche di GM12877 con frequenze prossime al 6,5% e al 13%, rispettivamente. È stato analizzato anche GM12877 non diluito. GM12877-13 è stato analizzato in duplicato a quattro livelli di input di DNA che andavano da 1000 ng a 25 ng (1000 ng, 250 ng, 50 ng e 25 ng). GM12877 è stato analizzato come singolo replicato a 250 ng e in duplicato per tutti gli altri livelli di input. Per la determinazione

dell'accuratezza, campioni di identificazione di varianti 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 presentata in [Risultati della PPA per ciascun input di DNA FFPE per tipo di variante e VAF target alla pagina 36](#). L'NPA è presentata in [NPA per ciascun input di DNA FFPE alla pagina 36](#). Tutti i livelli di input presentano un'accuratezza simile. Per i campioni FFPE con un valore ΔCq di ≤ 5 , l'input di DNA consigliato è di 50-1000 ng per il kit DNA Prep with Enrichment Dx di Illumina, in cui 1000 ng e 25 ng forniscono un limite superiore e inferiore per soddisfare le caratteristiche prestazionali quando vengono sequenziati su NovaSeq 6000Dx.

Tabella 11 Risultati della PPA per ciascun input di DNA FFPE per tipo di variante e VAF target

VAF diluizione target											
0,065											
0,13											
Input di DNA (ng)	Tipo di variante	Varianti previste	TP	FN	Identificazioni varianti non rilevate	PPA (%)	Varianti previste	TP	FN	Identificazioni varianti non rilevate	PPA (%)
25	SNV	3000	2931	8	61	99,7	624	624	0	0	100
50		3000	2930	8	62	99,7	624	622	0	2	100
250		3000	2927	8	65	99,7	624	624	0	0	100
1000		3000	2921	8	71	99,7	624	624	0	0	100
25	Inserzione	96	96	0	0	100	48	48	0	0	100
50		96	96	0	0	100	48	48	0	0	100
250		96	96	0	0	100	48	48	0	0	100
1000		96	96	0	0	100	48	48	0	0	100
25	Delezione	88	88	0	0	100	32	32	0	0	100
50		88	88	0	0	100	32	31	0	1	100
250		88	88	0	0	100	32	32	0	0	100
1000		88	88	0	0	100	32	32	0	0	100

Tabella 12 NPA per ciascun input di DNA FFPE

Input di DNA (ng)	Wild type previsto	TN	FP	Identificazioni riferimenti non rilevate	NPA (%)
25	25354119	25353706	413	5499498	>99,9
50	27538269	27538013	256	3315421	>99,9
250	21562303	21561983	320	1577958	>99,9
1000	29030903	29030596	307	1822781	>99,9

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

Questo studio è stato condotto per valutare il limite del bianco (LoB) e il limite del rilevamento (LoD) per il flusso di lavoro di analisi Somatic della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina su Strumento NovaSeq 6000Dx. Lo studio è stato eseguito utilizzando un test rappresentativo disegnato per interrogare una varietà di geni che coprono 1.970.505 basi su tutti i 23

cromosomi umani. Prima dell'estrazione del DNA, le linee cellulari GM12878 e GM12877 Platinum Genome sono state fissate in formalina e incluse in paraffina. Le diluizioni di GM12877 in GM12878 sono state preparate per ottenere campioni costituiti da 0%, 4%, 6,5% e 13% di GM12877 per volume, in modo tale che le frequenze delle varianti di 489 varianti univoche di GM12877 (454 SNV, 17 inserzioni e 18 delezioni) variassero tra 0 e 0,13. Le librerie di campioni sono state preparate utilizzando due lotti di reagenti del kit DNA Prep with Enrichment Dx di Illumina e sequenziate in sei giorni iniziali consecutivi con due Strumento NovaSeq 6000Dx e due lotti ciascuno di Kit reagenti NovaSeq 6000Dx S2 v1.5 (300 cicli) e Kit Reagenti NovaSeq 6000Dx S4 v1.5 (300 cicli), per un totale di dodici corse di sequenziamento. In questo modo sono state ottenute 288 osservazioni per ogni variante in ciascuna delle diluizioni del campione. LoB e LoD sono stati calcolati con l'approccio classico indicato da CLSI EP17-A2. LoB e LoD sono stati calcolati per i reagenti S2 e S4 separatamente, aggregando le frequenze di tutte le varianti nella corsa di sequenziamento per ciascun tipo di reagente. L'errore di tipo I era definito come 0,01 e l'errore di tipo II era definito come 0,05.

Il LoB è stato valutato per 489 loci in modo indipendente in due lotti di sequenziamento per ogni tipo di reagente (S2 o S4) e preparazione della libreria. Per i reagenti S2, il 95° percentile LoB era del 2,9%. Per i reagenti S4, il 95° percentile LoB era del 2,2%.

Il LoD è stato calcolato con successo per 478 delle 489 varianti per S2 e 485 delle 489 varianti per S4. Le varianti per le quali non è stata determinato alcun LoD per una o entrambe le preparazioni delle librerie sono state escluse dall'assegnazione finale del LoD per il sistema NovaSeq 6000Dx. Il LoD del sistema NovaSeq 6000Dx con i reagenti S2 e S4 è stato determinato prendendo il 95° percentile dei LoD delle singole varianti. Per i reagenti S2, il 95° percentile su 478 LoD delle varianti era del 4,8%. Per i reagenti S4, il 95° percentile su 485 LoD delle varianti era del 3,9%.

Accuratezza

Germline

Il seguente studio è stato condotto per valutare l'accuratezza dell'identificazione di varianti del flusso di lavoro di analisi Germline della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina su Strumento NovaSeq 6000Dx utilizzando Kit reagenti NovaSeq 6000Dx S2 v1.5 (300 cicli). Quattro campioni unici di DNA Platinum Genome sono stati analizzati con un test rappresentativo disegnato per interrogare una serie di geni che includono 1.970.505 basi (9.232 target) su tutti i 23 cromosomi umani. Ciascun campione è stato analizzato in repliche da 12, ad eccezione di NA12880, che è stato analizzato in repliche da 11. È stato eseguito un totale di 18 corse utilizzando tre strumenti di sequenziamento, tre lotti di reagenti S2 e due operatori su sei giorni di avvio. L'accuratezza è stata determinata per SNV, inserzioni e delezioni confrontando i risultati con la versione 2016-1.0 di Platinum Genomes.

Tabella 13 Riepilogo della concordanza per Germline

Criteri	Osservazioni totali ¹	Risultato dell'osservazione ²	Risultato per corsa ³
PPA per SNV	846	99,8	99,9
PPA per le inserzioni	846	97,9	>99,9
PPA per le delezioni	846	96,9	99,9
NPA	846	>99,9	>99,9
OPA	846	>99,9	>99,9

¹ Calcolato come il numero di campioni per corsa (47) x il numero di corse (18) = 846.

² Il valore più basso osservato per replicato di campione su tutte le 18 corse.

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

[Concordanza per campione per Germline alla pagina 39](#) (Concordanza per campione per Germline) 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 14 Concordanza per campione per Germline

Campione	Calcolabilità degli autosomi	Varianti previste ¹	TP	FN	Identificazioni varianti non rilevate	TN	FP	PPA	NPA	OPA
NA12877	99,4	273672	273452	220	0	414765131	931	99,9	>99,9	>99,9
NA12878	99,4	265680	265208	234	238	414803691	1193	99,9	>99,9	>99,9
NA12879	99,4	261792	261792	0	0	414746986	1429	100	>99,9	>99,9
NA12880	99,4	246114	245551	399	164	380157538	1458	99,8	>99,9	>99,9

¹ Il numero totale delle varianti in tutti i replicati dei campioni sulle 18 corse.

[Concordanza per campione in base al tipo di variante per Germline alla pagina 39](#) (Concordanza per campione in base al tipo di variante per Germline) 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 viene valutato per ogni tipo di variante - SNV, inserzioni e delezioni - separatamente. Sono state escluse le posizioni dei riferimenti.

Tabella 15 Concordanza per campione in base al tipo di variante per Germline

Campione	SVN			Inserzioni			Delezioni		
	Previsto	TP	FN	Previsto	TP	FN	Previsto	TP	FN
NA12877	255096	254877	219	10368	10367	1	8208	8208	0
NA12878	250344	250077	221	8424	8424	0	6912	6707	13
NA12879	246024	246024	0	8856	8856	0	6912	6912	0
NA12880	229482	229086	396	9306	9306	0	7326	7159	3

I campioni sono stati ulteriormente analizzati per identificare piccole inserzioni e delezioni (indel). Un riepilogo complessivo è presentato in [Riepilogo del rilevamento delle indel per Germline alla pagina 39](#). Era presente un totale di 210 indel con una dimensione compresa nell'intervallo 1–18 bp per le inserzioni e 1–21 bp per le delezioni.

Tabella 16 Riepilogo del rilevamento delle indel per Germline

Tipo di variante	Varianti previste	TP	FN	Identificazioni varianti non rilevate	PPA
Inserzione	36954	36953	1	0	>99,9
Delezione	29358	28986	16	356	99,9

Il test rappresentativo consisteva di 9.232 target che includevano una varietà di contenuto genomico. Il contenuto in GC dei target rientrava nell'intervallo compreso tra 0,20 e 0,86. I target presentavano inoltre un intervallo di ripetizioni di singolo nucleotide (ad es., PolyA, PolyT), dinucleotide e trinucleotide. I dati compilati su base cromosomica per determinare l'effetto del contenuto genomico sulla percentuale di chiamate corrette sono presentati in [Accuratezza a livello di cromosomi per Germline alla pagina 40](#). 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 cromosomi per Germline

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr1	47	728	138328	Poly A (12), Poly C (7), Poly T (14), Poly G (7), Dinucleotide (22), Trinucleotide (8), Inserzione (18), Delezione (4)	[0,22 - 0,8]; Mediana: 0,51	114888718	34	966860	>99,9	0,83

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr2	39	628	159588	Poly A (46), Poly C (8), Poly T (23), Poly G (7), Dinucleotide (22), Trinucleotide (8), Inserzione (5), Delezione (2)	[0,24 - 0,81]; Mediana: 0,44	132293464	798	460345	>99,9	0,35
chr3	38	650	137627	Poly A (18), Poly C (6), Poly T (18), Poly G (7), Dinucleotide (12), Trinucleotide (6), Inserzione (11), Delezione (1)	[0,25 - 0,86]; Mediana: 0,45	114625053	2	226461	>99,9	0,20

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr4	17	370	73766	Poly A (9), Poly C (7), Poly T (25), Poly G (6), Dinucleotide (5), Trinucleotide (5), Inserzione (2), Delezione (2)	[0,27 - 0,77]; Mediana: 0,45	61872303	0	66741	100	0,11
chr5	25	507	90008	Poly A (10), Poly C (6), Poly T (12), Poly G (7), Dinucleotide (10), Trinucleotide (8), Inserzione (8), Delezione (18)	[0,29 - 0,79]; Mediana: 0,46	75314497	912	153061	>99,9	0,20

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr6	39	453	126721	Poly A (28), Poly C (7), Poly T (33), Poly G (7), Dinucleotide (18), Trinucleotide (11), Inserzione (4), Delezione (2)	[0,24 - 0,79]; Mediana: 0,48	103412695	1	182361	>99,9	0,18
chr7	21	450	161501	Poly A (27), Poly C (8), Poly T (21), Poly G (7), Dinucleotide (31), Trinucleotide (5), Inserzione (1), Delezione (4)	[0,2 - 0,77]; Mediana: 0,46	132534074	19	246884	>99,9	0,19

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr8	18	381	67775	Poly A (19), Poly C (7), Poly T (13), Poly G (7), Dinucleotide (5), Trinucleotide (9), Inserzione (4), Delezione (1)	[0,26 - 0,78]; Mediana: 0,47	56247612	411	170925	>99,9	0,30
chr9	23	347	87100	Poly A (12), Poly C (7), Poly T (27), Poly G (8), Dinucleotide (9), Trinucleotide (9), Inserzione (4), Delezione (1)	[0,27 - 0,83]; Mediana: 0,49	72650800	20	241991	>99,9	0,33

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr10	14	317	66723	Poly A (26), Poly C (7), Poly T (15), Poly G (6), Dinucleotide (16), Trinucleotide (6), Inserzione (1), Delezione (1)	[0,23 - 0,78]; Mediana: 0,44	55539058	1	188216	>99,9	0,34
chr11	29	511	91786	Poly A (28), Poly C (8), Poly T (21), Poly G (7), Dinucleotide (26), Trinucleotide (7), Inserzione (2), Delezione (2)	[0,28 - 0,8]; Mediana: 0,47	75744222	742	259258	>99,9	0,34

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr12	29	577	120365	Poly A (19), Poly C (8), Poly T (40), Poly G (7), Dinucleotide (7), Trinucleotide (7), Inserzione (1), Delezione (5)	[0,26 - 0,77]; Mediana: 0,49	99972530	1	542005	>99,9	0,54
chr13	13	283	58639	Poly A (24), Poly C (6), Poly T (12), Poly G (7), Dinucleotide (6), Trinucleotide (8), Inserzione (14), Delezione (0)	[0,28 - 0,79]; Mediana: 0,42	48503179	1	45666	>99,9	0,09

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr14	11	147	26980	Poly A (21), Poly C (6), Poly T (18), Poly G (11), Dinucleotide (6), Trinucleotide (6), Inserzione (4), Delezione (1)	[0,29 - 0,77]; Mediana: 0,47	22286153	198	147895	>99,9	0,66
chr15	15	266	52091	Poly A (26), Poly C (7), Poly T (13), Poly G (6), Trinucleotide (8), Inserzione (4), Delezione (6)	[0,29 - 0,76]; Mediana: 0,46	43600279	0	99041	100	0,23

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr16	21	366	80030	Poly A (7), Poly C (7), Poly T (15), Poly G (7), Dinucleotide (5), Trinucleotide (10), Inserzione (15), Delezione (21)	[0,3 - 0,76]; Mediana: 0,54	65490245	16	1438278	>99,9	2,15
chr17	36	645	118062	Poly A (19), Poly C (7), Poly T (18), Poly G (8), Dinucleotide (13), Trinucleotide (6), Inserzione (18), Delezione (16)	[0,28 - 0,82]; Mediana: 0,49	97929929	417	335905	>99,9	0,34

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr18	9	99	19195	Poly A (7), Poly C (7), Poly T (15), Poly G (6), Trinucleotide (10), Inserzione (4), Delezione (0)	[0,22 - 0,78]; Mediana: 0,44	15967171	312	42077	>99,9	0,26
chr19	30	605	104004	Poly A (19), Poly C (7), Poly T (31), Poly G (7), Dinucleotide (5), Trinucleotide (7), Inserzione (2), Delezione (21)	[0,33 - 0,83]; Mediana: 0,59	85642066	3	678213	>99,9	0,79

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr20	12	179	33795	Poly A (6), Poly C (6), Poly T (7), Poly G (8), Trinucleotide (9), Inserzione (5), Delezione (0)	[0,31 - 0,84]; Mediana: 0,53	28108712	0	38374	100	0,14
chr21	5	63	30642	Poly A (28), Poly C (6), Poly T (24), Poly G (7), Dinucleotide (5), Inserzione (2), Delezione (5)	[0,22 - 0,78]; Mediana: 0,52	25319736	50	57434	>99,9	0,23

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr22	10	187	36727	Poly A (26), Poly C (7), Poly T (19), Poly G (7), Dinucleotide (5), Trinucleotide (6), Inserzione (6), Delezione (0)	[0,27 - 0,74]; Mediana: 0,51	30258131	0	42673	100	0,14
chrX	23	433	83576	Poly A (18), Poly C (8), Poly T (23), Poly G (9), Dinucleotide (5), Trinucleotide (23), Inserzione (3), Delezione (0)	[0,2 - 0,72]; Mediana: 0,48	67318722	0	770544	100	1,13
chrY	0	40	5476	Poly A (11), Poly C (8), Poly T (11), Poly G (5), Inserzione (0), Delezione (0)	[0,4 - 0,59]; Mediana: 0,45	0	0	0	N/D	N/D

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 9.232 target, 8.009 sono rientrati completamente nelle regioni genomiche altamente affidabili, 776 presentavano una sovrapposizione parziale e 447 non presentavano alcuna sovrapposizione nella sequenza NIST. Questo ha fornito 1.831.483 coordinate per replicato per il confronto. Le identificazioni delle basi non varianti sono state confrontate con la sequenza di riferimento del genoma umano versione hg19. I risultati dell'accuratezza sono mostrati in [Concordanza del campione NA12878 rispetto al database NIST per Germline alla pagina 52](#).

Tabella 18 Concordanza del campione NA12878 rispetto al database NIST per Germline

Campione	N. di target inclusi	Calcolabilità degli autosomi	TP	FN	TN	FP	PPA	NPA	OPA
NA12878	8785	99,4	247709	218	394262149	4584	>99,9	>99,9	>99,9

In base ai dati forniti da questo studio Germline di 18 corse, Strumento NovaSeq 6000Dx può sequenziare in modo coerente:

- Contenuto in GC \geq 20% (tutte le basi identificate in 1.692 regioni target sequenziate con 20% di contenuto in GC identificato correttamente con una percentuale di identificazioni non rilevate di 0%)
- Contenuto in GC \leq 86% (tutte le basi identificate in 846 regioni target sequenziate con 86% di contenuto in GC identificato correttamente con una percentuale di identificazioni non rilevate di 0%)
- Lunghezze PolyA \leq 46 (tutte le basi identificate in 846 regioni target sequenziate contenenti 46 ripetizioni PolyA identificate correttamente con una percentuale di identificazioni non rilevate di 0,27%)
- Lunghezze PolyT \leq 40 (13.384.074 su 13.384.321 basi identificate in 846 regioni target sequenziate contenenti 40 ripetizioni PolyT identificate correttamente con una percentuale di identificazioni non rilevate di 0,26%)
- Lunghezze PolyG \leq 11 (tutte le basi identificate in 846 regioni target sequenziate contenenti 11 ripetizioni PolyG identificate correttamente con una percentuale di identificazioni non rilevate di 0%)
- Lunghezze PolyC \leq 8 (9.815.030 su 9.815.035 basi identificate in 5.922 regioni target sequenziate contenenti 8 ripetizioni PolyC identificate correttamente con una percentuale di identificazioni non rilevate di 0,53%)
- Lunghezze di dinucleotidi ripetuti \leq 31x (32.233.922 su 32.233.926 basi identificate in 846 regioni target sequenziate con 31 ripetizioni di dinucleotidi identificate correttamente con una percentuale di identificazioni non rilevate dello 0,21%)
- Lunghezze di trinucleotidi ripetuti \leq 23x (tutte le basi identificate in 846 delle regioni target sequenziate con 23 ripetizioni di trinucleotidi identificate correttamente con una percentuale di identificazioni non rilevate dello 0,21%)

- Lunghezze di inserzione ≤ 18 (tutte le basi identificate in 1.692 regioni target sequenziate con 18 inserzioni identificate correttamente con una percentuale di identificazioni non rilevate di 7,71%)
- Lunghezze di delezione ≤ 21 (tutte le basi identificate in 1.692 regioni target sequenziate con 21 delezioni identificate correttamente con una percentuale di identificazioni non rilevate di 1,14%)

Somatic

Lo studio qui descritto è stato utilizzato per valutare l'accuratezza dell'identificazione di varianti del flusso di lavoro di analisi Somatic della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina su Strumento NovaSeq 6000Dx utilizzando Kit Reagenti NovaSeq 6000Dx S4 v1.5 (300 cicli).

Questo studio ha utilizzato un test rappresentativo disegnato per interrogare una serie di geni che coprono 1.970.505 basi (9.232 target) su tutti i 23 cromosomi umani. Nello studio, il DNA Platinum Genome è stato estratto da blocchetti trattati in FFPE per generare quattro campioni univoci da utilizzare per la valutazione.

Il campione di DNA GM12877 è stato diluito con il campione di DNA GM12878 per creare GM12877-13 con varianti eterozigoti e omozigoti univoche di GM12877 con frequenze prossime al 6,5% e al 13%, rispettivamente. Il campione di DNA GM12878 è stato diluito in modo simile con il campione di DNA GM12877 per creare GM12878-13 con varianti eterozigoti e omozigoti univoche di GM12878 con frequenze prossime al 6,5% e al 13%, rispettivamente. Sono stati testati anche GM12877 e GM12878 non diluiti. Ciascun campione è stato analizzato in replicati da 12, tranne GM12878 non diluito, che è stato analizzato in replicati da undici. È stato eseguito un totale di diciotto corse utilizzando tre strumenti di sequenziamento, tre lotti di reagenti S4 e due operatori su sei giorni di avvio. L'accuratezza è stata determinata per SNV, inserzioni e delezioni confrontando i risultati con la versione 2016-1.0 di Platinum Genomes.

Tabella 19 Riepilogo della concordanza Somatic

Criteri	N. di osservazioni ¹	Risultato delle osservazioni ²	Risultato per corsa ³
PPA per SNV somatici	846	99,8	98,9
PPA per inserzioni somatiche	846	100	100
PPA per delezioni somatiche	846	100	100
NPA	846	>99,9	>99,9
OPA	846	>99,9	>99,9

¹ Calcolato come = N. di campioni per corsa (47) x N. di corse (18) = 846.

² Valore più basso osservato per replicato di campione su tutte le 18 corse.

³ Valore più basso quando i dati di ciascuna corsa vengono analizzati in aggregato.

[Concordanza per campione per Somatic alla pagina 54](#) 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 20 Concordanza per campione per Somatic

Campione	Calcolabilità degli autosomi	Varianti previste	TP	FN	Identificazioni varianti non rilevate	TN	FP	PPA	NPA	OPA
GM12877	95,4	96228	95022	198	1008	365425810	1203	99,8	>99,9	>99,9
GM12878	94,5	96768	96278	0	490	395002023	1278	100	>99,9	>99,9
GM12877-13	94,7	104976	103029	216	1731	395989324	1286	99,8	>99,9	>99,9
GM12878-13	95,2	96768	96027	0	741	397900884	1218	100	>99,9	>99,9

[Concordanza per campione in base al tipo di variante per Somatic alla pagina 54](#) 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 viene valutato per ogni tipo di variante - SNV, inserzioni e delezioni - separatamente. Sono state escluse le posizioni dei riferimenti.

Tabella 21 Concordanza per campione in base al tipo di variante per Somatic

Campione	SNV			Inserzioni			Delezioni		
	Previsto	TP	FN	Previsto	TP	FN	Previsto	TP	FN
GM12877	89694	88488	198	3564	3564	0	2970	2970	0
GM12878	92664	92390	0	2160	2160	0	1944	1728	0
GM12877-13	97848	95901	216	3888	3888	0	3240	3240	0
GM12878-13	92664	92139	0	2160	2160	0	1944	1728	0

I quattro campioni sono stati ulteriormente analizzati per identificare piccole inserzioni e delezioni (indel). Una sintesi generale è presentata in [Riepilogo del rilevamento delle indel per Somatic alla pagina 55](#). Era presente un totale di 210 indel con una dimensione compresa nell'intervallo 1–18 bp per le inserzioni e 1–21 bp per le delezioni.

Tabella 22 Riepilogo del rilevamento delle indel per Somatic

Tipo di variante	Varianti previste	TP	FN	Identificazioni varianti non rilevate	PPA
Inserzione	11772	11772	0	0	100
Delezione	10098	9666	0	432	100

Il test rappresentativo consisteva di 9.232 target che includevano una varietà di contenuto genomico. Il contenuto in GC dei target rientrava nell'intervallo compreso tra 0,20 e 0,86. I target presentavano inoltre un intervallo di ripetizioni di singolo nucleotide (ad es., PolyA, PolyT), dinucleotide e trinucleotide. I dati compilati su base cromosomica per determinare l'effetto del contenuto genomico sulla percentuale di chiamate corrette sono presentati in [Accuratezza a livello di cromosomi per Somatic alla pagina 55](#). 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 23 Accuratezza a livello di cromosomi per Somatic

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr1	47	728	138328	Poly A (12), Poly C (7), Poly T (14), Poly G (7), Dinucleotide (22), Trinucleotide (8), Inserzione (3), Delezione (0)	[0,22 - 0,8]; Mediana: 0,51	110145939	52	5642613	>99,9	4,9

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr2	39	628	159588	Poly A (46), Poly C (8), Poly T (23), Poly G (7), Dinucleotide (22), Trinucleotide (8), Inserzione (5), Delezione (1)	[0,24 - 0,81]; Mediana: 0,44	126795713	842	5850393	>99,9	4,4
chr3	38	650	137627	Poly A (18), Poly C (6), Poly T (18), Poly G (7), Dinucleotide (12), Trinucleotide (6), Inserzione (1), Delezione (1)	[0,25 - 0,86]; Mediana: 0,45	109902527	593	4889226	>99,9	4,3

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr4	17	370	73766	Poly A (9), Poly C (7), Poly T (25), Poly G (6), Dinucleotide (5), Trinucleotide (5), Inserzione (0), Delezione (1)	[0,27 - 0,77]; Mediana: 0,45	59373461	16	2517412	>99,9	4,1
chr5	25	507	90008	Poly A (10), Poly C (6), Poly T (12), Poly G (7), Dinucleotide (10), Trinucleotide (8), Inserzione (8), Delezione (18)	[0,29 - 0,79]; Mediana: 0,46	72261191	723	3116981	>99,9	4,1

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr6	39	453	126721	Poly A (28), Poly C (7), Poly T (33), Poly G (7), Dinucleotide (18), Trinucleotide (11), Inserzione (0), Delezione (1)	[0,24 - 0,79]; Mediana: 0,48	98593101	687	4890221	>99,9	4,7
chr7	21	450	161501	Poly A (27), Poly C (8), Poly T (21), Poly G (7), Dinucleotide (31), Trinucleotide (5), Inserzione (1), Delezione (4)	[0,2 - 0,77]; Mediana: 0,46	126913574	104	5773856	>99,9	4,4

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr8	18	381	67775	Poly A (19), Poly C (7), Poly T (13), Poly G (7), Dinucleotide (5), Trinucleotide (9), Inserzione (4), Delezione (0)	[0,26 - 0,78]; Mediana: 0,47	53430489	175	2958909	>99,9	5,2
chr9	23	347	87100	Poly A (12), Poly C (7), Poly T (27), Poly G (8), Dinucleotide (9), Trinucleotide (9), Inserzione (0), Delezione (1)	[0,27 - 0,83]; Mediana: 0,49	69594586	74	3260257	>99,9	4,5

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr10	14	317	66723	Poly A (26), Poly C (7), Poly T (15), Poly G (6), Dinucleotide (16), Trinucleotide (6), Inserzione (0), Delezione (0)	[0,23 - 0,78]; Mediana: 0,44	53209592	90	2469444	>99,9	4,4
chr11	29	511	91786	Poly A (28), Poly C (8), Poly T (21), Poly G (7), Dinucleotide (26), Trinucleotide (7), Inserzione (2), Delezione (2)	[0,28 - 0,8]; Mediana: 0,47	72291795	150	3665560	>99,9	4,8

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr12	29	577	120365	Poly A (19), Poly C (8), Poly T (40), Poly G (7), Dinucleotide (7), Trinucleotide (7), Inserzione (0), Delezione (3)	[0,26 - 0,77]; Mediana: 0,49	96109352	101	4331932	>99,9	4,3
chr13	13	283	58639	Poly A (24), Poly C (6), Poly T (12), Poly G (7), Dinucleotide (6), Trinucleotide (8), Inserzione (14), Delezione (0)	[0,28 - 0,79]; Mediana: 0,42	46130028	44	2384839	>99,9	4,9

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr14	11	147	26980	Poly A (21), Poly C (6), Poly T (18), Poly G (11), Dinucleotide (6), Trinucleotide (6), Inserzione (4), Delezione (0)	[0,29 - 0,77]; Mediana: 0,47	21336891	0	1078329	100	4,8
chr15	15	266	52091	Poly A (26), Poly C (7), Poly T (13), Poly G (6), Trinucleotide (8), Inserzione (4), Delezione (0)	[0,29 - 0,76]; Mediana: 0,46	41918631	184	1753300	>99,9	4,0

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr16	21	366	80030	Poly A (7), Poly C (7), Poly T (15), Poly G (7), Dinucleotide (5), Trinucleotide (10), Inserzione (15), Delezione (21)	[0,3 - 0,76]; Mediana: 0,54	62344351	18	4540539	>99,9	6,8
chr17	36	645	118062	Poly A (19), Poly C (7), Poly T (18), Poly G (8), Dinucleotide (13), Trinucleotide (6), Inserzione (18), Delezione (1)	[0,28 - 0,82]; Mediana: 0,49	93811318	414	4403622	>99,9	4,5

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr18	9	99	19195	Poly A (7), Poly C (7), Poly T (15), Poly G (6), Trinucleotide (10), Inserzione (0), Delezione (0)	[0,22 - 0,78]; Mediana: 0,44	15007653	6	990633	>99,9	6,2
chr19	30	605	104004	Poly A (19), Poly C (7), Poly T (31), Poly G (7), Dinucleotide (5), Trinucleotide (7), Inserzione (2), Delezione (3)	[0,33 - 0,83]; Mediana: 0,59	81416722	455	4860311	>99,9	5,6

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr20	12	179	33795	Poly A (6), Poly C (6), Poly T (7), Poly G (8), Trinucleotide (9), Inserzione (5), Delezione (0)	[0,31 - 0,84]; Mediana: 0,53	26833936	7	1301905	>99,9	4,6
chr21	5	63	30642	Poly A (28), Poly C (6), Poly T (24), Poly G (7), Dinucleotide (5), Inserzione (1), Delezione (0)	[0,22 - 0,78]; Mediana: 0,52	24169250	44	1172087	>99,9	4,6

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chr22	10	187	36727	Poly A (26), Poly C (7), Poly T (19), Poly G (7), Dinucleotide (5), Trinucleotide (6), Inserzione (6), Delezione (0)	[0,27 - 0,74]; Mediana: 0,51	28887217	86	1392179	>99,9	4,6
chrX	23	433	83576	Poly A (18), Poly C (8), Poly T (23), Poly G (9), Dinucleotide (5), Trinucleotide (23), Inserzione (3), Delezione (0)	[0,2 - 0,72]; Mediana: 0,48	64231080	241	3852253	>99,9	5,7

Cromosoma	N. geni	N. target	N. basi	Contenuto genomico	Gamma GC	Identificazioni corrette	Identificazioni errate	Nessuna identificazione	% di identificazioni corrette	% di identificazioni non riuscite
chrY	0	40	5476	Poly A (11), Poly C (8), Poly T (11), Poly G (5), Inserzione (0), Delezione (0)	[0,4 - 0,59]; Mediana: 0,45	0	0	0	N/D	N/D

I risultati del sequenziamento per il campione GM12878 sono stati confrontati con un genotipo altamente affidabile per NA12878, stabilito dal National Institute of Standards and Technology (NIST) (v.2.19). Dei 9.232 target, 8.009 sono rientrati completamente nelle regioni genomiche altamente affidabili, 776 presentavano una sovrapposizione parziale e 447 non presentavano alcuna sovrapposizione nella sequenza NIST. Questo ha fornito 1.831.483 coordinate per replicato per il confronto. Le identificazioni delle basi non varianti sono state confrontate con la sequenza di riferimento del genoma umano versione hg19. I risultati di accuratezza sono mostrati in [Concordanza del campione GM12878 rispetto al database NIST per Somatic alla pagina 67](#).

Tabella 24 Concordanza del campione GM12878 rispetto al database NIST per Somatic

Campione	N. di target inclusi	Calcolabilità degli autosomi	TP	FN	TN	FP	PPA	NPA	OPA
GM12878	8785	94,5	247228	0	375073821	2043	100	>99,9	>99,9

In base ai dati forniti da questo studio Somatic di 18 corse, Strumento NovaSeq 6000DX può sequenziare in modo coerente:

- Contenuto in GC \geq 20% (tutte le basi identificate in 1.692 regioni target sequenziate con 20% di contenuto in GC identificato correttamente con una percentuale di identificazioni non rilevate di 0,34%)
- Contenuto in GC \leq 86% (tutte le basi identificate in 846 regioni target sequenziate con l'86% di contenuto in GC identificato correttamente con una percentuale di identificazioni non rilevate del 4,21%)
- Lunghezze PolyA \leq 46 (14.550.082 basi identificate su 14.550.083 in 846 regioni target sequenziate contenenti 46 ripetizioni PolyA identificate correttamente con una percentuale di identificazioni non rilevate del 4,18%)
- Lunghezze PolyT \leq 40 (12.833.489 basi identificate su 12.833.491 in 846 regioni target sequenziate contenenti 40 ripetizioni PolyT identificate correttamente con una percentuale di identificazioni non rilevate del 4,37%)

- Lunghezze PolyG ≤ 11 (tutte le basi identificate in 846 regioni target sequenziate contenenti 11 ripetizioni PolyG identificate correttamente con una percentuale di identificazioni non rilevate del 7,59%)
- Lunghezze PolyC ≤ 8 (9.405.604 su 9.405.615 basi identificate in 5.922 regioni target sequenziate contenenti 8 ripetizioni PolyC identificate correttamente con una percentuale di identificazioni non rilevate del 4,68%)
- Lunghezze di dinucleotidi ripetuti $\leq 31x$ (30.996.684 basi identificate su 30.996.712 in 846 regioni target sequenziate contenenti 31 ripetizioni di dinucleotidi identificate correttamente con una percentuale di identificazioni non rilevate del 4,04%)
- Lunghezze di trinucleotidi ripetuti $\leq 23x$ (tutte le basi identificate in 846 regioni target sequenziate contenenti 23 ripetizioni di trinucleotidi identificate correttamente con una percentuale di identificazioni non rilevate del 5,39%)
- Lunghezze inserzioni ≤ 18 (tutte le basi identificate in 846 regioni target sequenziate contenenti 18 inserzioni identificate correttamente con una percentuale di identificazioni non rilevate di 1,44%)
- Lunghezze di delezione ≤ 21 (tutte le basi identificate in 846 regioni target sequenziate contenenti 21 delezioni identificate correttamente con una percentuale di identificazioni non rilevate di 7,86%)

Precisione

La precisione di Strumento NovaSeq 6000Dx è stata valutata utilizzando i campioni Platinum Genome con un saggio rappresentativo progettato per interrogare diversi geni che coprono 1.970.505 basi su 23 diversi cromosomi utilizzando 9.232 oligo target. Sono state valutate 1.723 piccole varianti mirate (SNV, inserzioni e delezioni). Le analisi Germline consistevano in undici o dodici repliche di quattro campioni univoci Platinum Genome. Le analisi Somatic consistevano in undici o dodici repliche di quattro campioni univoci Platinum Genome trattati con FFPE a diversi livelli di VAF. Le librerie di campioni sono state preparate utilizzando i reagenti del kit DNA Prep with Enrichment Dx di Illumina.

Le analisi sono state eseguite in un sito interno utilizzando tre Strumento NovaSeq 6000Dx, tre lotti ciascuno di Kit reagenti NovaSeq 6000Dx S2 v1.5 (300 cicli) e Kit Reagenti NovaSeq 6000Dx S4 v1.5 (300 cicli), e due operatori per sei giorni di avvio. Per ogni giorno di avvio, le librerie di campioni di linea germinale sono state sequenziate su un lato dello strumento utilizzando i reagenti S2 e il flusso di lavoro di analisi Germline della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina, mentre le librerie di campioni somatici sono state sequenziate sull'altro lato dello strumento utilizzando i reagenti S4 e il flusso di lavoro di analisi Somatic della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina. Questo ha fornito 18 celle a flusso per ciascun flusso di lavoro per le linee germinali e per le varianti somatiche.

Germline

Per le corse della linea germinale, le posizioni genomiche in cui viene rilevata una variante della linea germinale mirata sono segnalate 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 (PPC) entro ciascun tipo di variante (SNV, inserzioni, delezioni). [Osservazioni di precisione in laboratorio delle identificazioni per i risultati positivi previsti in base al tipo di variante per Germline alla pagina 69](#) riassume le percentuali osservate, assieme al livello superiore e inferiore di affidabilità (Lower Confidence Limit (LCL)/Upper Confidence Limit (UCL)) del 95% calcolato in base al metodo del punteggio di Wilson Score, per ciascun tipo di variante.

Tabella 25 Osservazioni di precisione in laboratorio delle identificazioni per i risultati positivi previsti in base al tipo di variante per Germline

Tipo di variante	Identificazioni non riuscite osservate ¹	Identificazioni totali	Percentuale di identificazioni non riuscite	Identificazioni positive osservate ²	Identificazioni totali valutabili	PPC	LCL al 95% ³	UCL 95%
SNV	6	980316	<0,01	979854	980310	99,95	99,95	99,96
Inserzione	0	36738	0	36738	36738	100	>99,99	100
Delezione	18	34434	0,05	32160	34416	93,44	93,18	93,70

¹ Identificazione non riuscita definita come posizione cromosomica mirata in cui non è possibile determinare una variante (a causa della bassa profondità di copertura).

² Identificazione riuscita definita come posizioni cromosomiche mirate in cui viene rilevata una variante.

³ Gli intervalli di confidenza bilaterali al 95% sono calcolati utilizzando il metodo del punteggio di Wilson.

Le posizioni genomiche in cui non viene rilevata una variante mirata sono riportate come negative (wild type). Per le posizioni negative previste, i dati sono stati valutati per le identificazioni non riuscite e le percentuali di identificazioni negative (PNC). [Osservazioni di precisione in laboratorio delle identificazioni per i risultati negativi previsti alla pagina 70](#) riassume le percentuali osservate, assieme al livello superiore e inferiore di affidabilità (Lower Confidence Limit (LCL)/Upper Confidence Limit (UCL)) del 95% calcolato in base al metodo del punteggio di Wilson.

Tabella 26 Osservazioni di precisione in laboratorio delle identificazioni per i risultati negativi previsti

Tipo di variante	Identificazioni non riuscite osservate ¹	Identificazioni totali	Percentuale di identificazioni non riuscite	Identificazioni negative osservate ²	Identificazioni totali valutabili	PNC	LCL al 95% ³	UCL 95%
Wild type	0	406170	0	406170	406170	100	>99,99	100

¹ Identificazione non riuscita definita come posizione cromosomica mirata in cui non è possibile determinare una variante (a causa della bassa profondità di copertura).

² Identificazione negativa definita come posizioni cromosomiche mirate in cui non viene rilevata una variante.

³ Gli intervalli di confidenza bilaterali al 95% sono calcolati utilizzando il metodo del punteggio di Wilson.

Il contributo di ciascun parametro (strumento, lotto di reagente, giorno, replica della libreria) alla variabilità complessiva è stato determinato mediante l'analisi delle componenti della varianza utilizzando la frequenza delle varianti come variabile di risposta. La deviazione standard complessiva aveva una media di 0,0370. Il principale fattore contribuente alla variabilità della frequenza delle varianti proveniva dalle repliche della preparazione della libreria, che hanno contribuito al 17,1% della variabilità complessiva. Il giorno ha contribuito all'1%, mentre lo strumento e il lotto del reagente hanno contribuito ciascuno a meno dell'1% della variabilità totale. [Stime delle componenti della varianza della precisione all'interno del laboratorio per le frequenze delle varianti dei campioni per Germline alla pagina 70](#) (DS = deviazione standard).

Tabella 27 Stime delle componenti della varianza della precisione all'interno del laboratorio per le frequenze delle varianti dei campioni per Germline

Componente	DS media	% media della DS totale
Giorno	0,0020	1,028
Strumento	0,0018	0,837
Lotto di materiali di consumo	0,0016	0,712
Replicato delle librerie	0,0143	17,110
Totale	0,0370	100

Somatic

Per le corse somatiche, le posizioni genomiche in cui viene rilevata una variante somatica mirata sono segnalate come positive (variante). Per i campioni diluiti GM12877-13 e GM12878-13 con varianti somatiche positive previste a VAF compresi tra il 6,5% e il 13%, i dati sono stati valutati per il tasso di identificazioni non riuscite e la

percentuale di identificazioni positive (PPC) all'interno di ciascun tipo di variante (SNV, inserzione, delezione). [Osservazione delle identificazioni di precisione all'interno del laboratorio per i risultati positivi previsti in base al tipo di variante per Somatic \(la VAF è \$\geq 6,5\%\$ e \$\leq 13\%\$ \) alla pagina 71](#) riassume le percentuali osservate, insieme ai livelli di confidenza al 95% inferiori e superiori (LCL/UCL) calcolati con il metodo del punteggio di Wilson, per ciascun tipo di variante.

Tabella 28 Osservazione delle identificazioni di precisione all'interno del laboratorio per i risultati positivi previsti in base al tipo di variante per Somatic (la VAF è $\geq 6,5\%$ e $\leq 13\%$)

Tipo di variante	Identificazioni non riuscite osservate ¹	Identificazioni totali	Percentuale di identificazioni non riuscite	Identificazioni positive osservate ²	Identificazioni totali valutabili	PPC	LCL al 95% ³	UCL 95%
SNV	0	96939	0	96069	96939	99,10	99,04	99,16
Inserzione	0	3004	0	3004	3004	100	99,87	100
Delezione	0	2912	0	2907	2912	99,83	99,60	99,93

¹ Identificazione non riuscita definita come posizione cromosomica mirata in cui non è possibile determinare una variante (a causa della bassa profondità di copertura).

² Identificazione riuscita definita come posizioni cromosomiche mirate in cui viene rilevata una variante.

³ Gli intervalli di confidenza bilaterali al 95% sono calcolati utilizzando il metodo del punteggio di Wilson.

Le posizioni genomiche in cui non viene rilevata una variante somatica mirata sono riportate come negative (wild type). Per le posizioni negative previste, i dati sono stati valutati per la percentuale di identificazioni non rilevate e la percentuale di identificazioni negative. [Osservazioni delle identificazioni di precisione all'interno del laboratorio per i risultati negativi previsti per Somatic alla pagina 71](#) riassume le percentuali osservate, assieme al livello superiore e inferiore di affidabilità (Lower Confidence Limit (LCL)/Upper Confidence Limit (UCL)) del 95% calcolato in base al metodo del punteggio di Wilson Score, per ciascun tipo di variante.

Tabella 29 Osservazioni delle identificazioni di precisione all'interno del laboratorio per i risultati negativi previsti per Somatic

Tipo di variante	Identificazioni non riuscite osservate ¹	Identificazioni totali	Percentuale di identificazioni non riuscite	Identificazioni negative osservate ²	Identificazioni totali valutabili	PNC	LCL al 95% ³	UCL 95%
Wild type	0	194922	0	194919	194922	>99,99	>99,99	100

¹ Identificazione non riuscita definita come posizione cromosomica mirata in cui non è possibile determinare una variante (a causa della bassa profondità di copertura).

² Identificazione negativa definita come posizioni cromosomiche mirate in cui non viene rilevata una variante.

³ Gli intervalli di confidenza bilaterali al 95% sono calcolati utilizzando il metodo del punteggio di Wilson.

Il contributo di ciascun parametro (strumento, lotto di reagente, giorno, replica della libreria) alla variabilità complessiva è stato determinato mediante l'analisi delle componenti della varianza utilizzando la frequenza delle varianti come variabile di risposta. La deviazione standard complessiva aveva una media di 0,0062. Le repliche della preparazione della libreria sono rimaste la fonte di variabilità più significativa, pari al 50,7% del

totale. Il giorno, lo strumento e il lotto di materiali di consumo hanno contribuito a meno dell'1% della variabilità totale. [Stime delle componenti della varianza della precisione all'interno del laboratorio per le frequenze delle varianti dei campioni per Somatic alla pagina 72](#) (DS = deviazione standard).

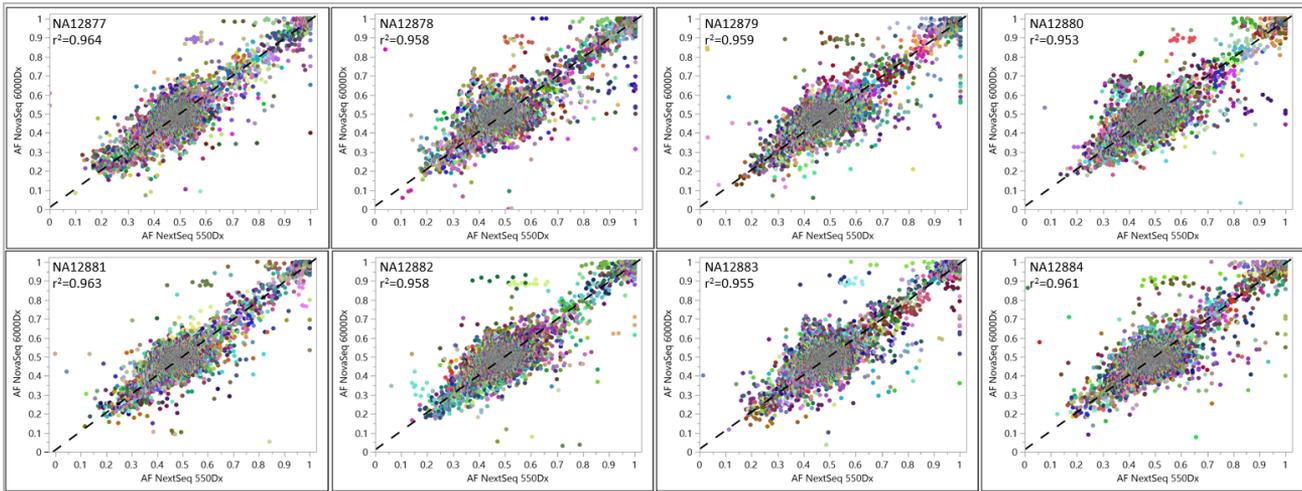
Tabella 30 Stime delle componenti della varianza della precisione all'interno del laboratorio per le frequenze delle varianti dei campioni per Somatic

Componente	DS media	% media della DS totale
Giorno	0,0002	0,41
Strumento	0,0002	0,40
Lotto di materiali di consumo	0,0002	0,35
Replicato delle librerie	0,0044	50,7
Totale	0,0062	100

Confronto del metodo

È stato condotto uno studio per confrontare le prestazioni tra gli strumenti NovaSeq 6000Dx e NextSeq 550Dx. La concordanza sulla frequenza delle varianti per i campioni di sangue è stata valutata utilizzando un test rappresentativo progettato per interrogare una varietà di geni che coprono 1.970.505 basi su tutti i 23 cromosomi umani. Sono stati analizzati otto campioni di DNA Platinum Genome, sette in repliche di sei e uno (NA12881) in repliche di cinque. Le librerie sono state sequenziate su Strumento NovaSeq 6000Dx utilizzando il flusso di lavoro di analisi Germline della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina e sullo strumento NextSeq 550Dx utilizzando il modulo DNA Generate FASTQ Dx Local Run Manager. [Grafici di correlazione della frequenza delle varianti \(i punti sono colorati in base alla variante univoca. Le varianti possono essere colorate in modo diverso in ogni singolo grafico\).](#) alla pagina 73 traccia la correlazione VAF tra i due strumenti per ciascun campione. Sulla base della forte correlazione tra Strumento NovaSeq 6000Dx 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, consultare il foglio illustrativo di DNA Prep with Enrichment Dx di Illumina.

Figura 15 Grafici di correlazione della frequenza delle varianti (i punti sono colorati in base alla variante univoca. Le varianti possono essere colorate in modo diverso in ogni singolo grafico).



Riproducibilità

La riproducibilità di Strumento NovaSeq 6000Dx è stata valutata utilizzando i campioni Platinum Genome con un saggio rappresentativo progettato per interrogare diversi geni che coprono 1.970.505 basi su 23 diversi cromosomi utilizzando 9.232 oligo target. Sono state valutate 1.723 piccole varianti mirate (SNV, inserzioni e delezioni). L'analisi Germline consisteva in tre o quattro repliche di dodici campioni Platinum univoci. L'analisi Somatic consisteva in cinque o sei repliche di otto campioni Platinum Genome univoci trattati con FFPE a diversi livelli di VAF. Le librerie di campioni sono state preparate utilizzando i reagenti del kit DNA Prep with Enrichment Dx di Illumina.

L'analisi è stata eseguita in tre siti esterni utilizzando un lotto ciascuno di Kit reagenti NovaSeq 6000Dx S2 v1.5 (300 cicli) e Kit Reagenti NovaSeq 6000Dx S4 v1.5 (300 cicli). Un singolo Strumento NovaSeq 6000Dx è stato utilizzato in ogni sito. 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 celle a flusso sui tre siti. Per ogni giorno di avvio, le librerie di campioni della linea germinale sono state sequenziate sul lato A dello strumento utilizzando i reagenti S2 e il flusso di lavoro di analisi Germline della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina, mentre le librerie di campioni somatici sono state sequenziate sul lato B dello strumento utilizzando i reagenti S4 e il flusso di lavoro di analisi Somatic della generazione di file FASTQ e VCF dell'applicazione DRAGEN per DNA Prep with Enrichment Dx di Illumina. Questo ha fornito 18 celle a flusso per ciascun flusso di lavoro per le linee germinali e per le varianti somatiche.

Germline

Per le corse della linea germinale, le posizioni genomiche in cui viene rilevata una variante della linea germinale mirata sono segnalate 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 (PPC) entro ciascun tipo di variante (SNV, inserzioni, delezioni). [Osservazioni delle identificazioni per i risultati positivi](#)

[previsti in base al tipo di variante per Germline alla pagina 74](#) riassume le percentuali osservate, assieme al livello superiore e inferiore di affidabilità (Lower Confidence Limit (LCL)/Upper Confidence Limit (UCL)) del 95% calcolato in base al metodo del punteggio di Wilson Score, per ciascun tipo di variante.

Tabella 31 Osservazioni delle identificazioni per i risultati positivi previsti in base al tipo di variante per Germline

Tipo di variante	Identificazioni non riuscite osservate ¹	Identificazioni totali	Percentuale di identificazioni non riuscite	Identificazioni positive osservate ²	Identificazioni totali valutabili	PPC	LCL al 95% ³	UCL 95%
SNV	0	991026	0	990276	991026	99,92	99,92	99,93
Inserzione	0	38358	0	38358	38358	100	99,99	100
Delezione	0	34758	0	32228	34758	92,72	92,44	92,99

¹ Identificazione non riuscita definita come posizione cromosomica mirata in cui non è possibile determinare una variante (a causa della bassa profondità di copertura).

² Identificazione riuscita definita come posizioni cromosomiche mirate in cui viene rilevata una variante.

³ Gli intervalli di confidenza bilaterali al 95% sono calcolati utilizzando il metodo del punteggio di Wilson.

Le posizioni genomiche in cui non viene rilevata una variante mirata sono riportate come negative (wild type). Per le posizioni negative previste, i dati sono stati valutati per le identificazioni non riuscite e le percentuali di identificazioni negative (PNC). [Osservazione delle identificazioni per i risultati negativi previsti per Germline alla pagina 74](#) riassume le percentuali osservate, assieme al livello superiore e inferiore di affidabilità (Lower Confidence Limit (LCL)/Upper Confidence Limit (UCL)) del 95% calcolato in base al metodo del punteggio di Wilson.

Tabella 32 Osservazione delle identificazioni per i risultati negativi previsti per Germline

Tipo di variante	Identificazioni non riuscite osservate ¹	Identificazioni totali	Percentuale di identificazioni non riuscite	Identificazioni negative osservate ²	Identificazioni totali valutabili	PNC	LCL al 95% ³	UCL 95%
Wild type	0	393516	0	393516	393516	100	>99,99	100

¹ Identificazione non riuscita definita come posizione cromosomica mirata in cui non è possibile determinare una variante (a causa della bassa profondità di copertura).

² Identificazione negativa definita come posizioni cromosomiche mirate in cui non viene rilevata una variante.

³ Gli intervalli di confidenza bilaterali al 95% sono calcolati utilizzando il metodo del punteggio di Wilson.

Somatic

Per le corse somatiche, le posizioni genomiche in cui viene rilevata una variante somatica mirata sono segnalate come positive (variante). Per le varianti somatiche positive previste in cui la frequenza allelica media della variante (Variant Allele Frequency, VAF) è maggiore o pari al 14% e minore o pari al 28%, i dati sono stati valutati per il tasso di mancata identificazione e per la percentuale di identificazioni positive (PPC) all'interno di ciascun tipo di variante (SNV, inserzione, delezione). [Osservazioni delle identificazioni per i risultati positivi previsti in](#)

base al tipo di variante (VAF media $\geq 14\%$ e $\leq 28\%$) per Somatic alla pagina 75 riassume le percentuali osservate, assieme al livello superiore e inferiore di affidabilità (Lower Confidence Limit (LCL)/Upper Confidence Limit (UCL)) del 95% calcolato in base al metodo del punteggio di Wilson, per ciascun tipo di variante.

Tabella 33 Osservazioni delle identificazioni per i risultati positivi previsti in base al tipo di variante (VAF media $\geq 14\%$ e $\leq 28\%$) per Somatic

Tipo di variante	Identificazioni non riuscite osservate ¹	Identificazioni totali	Percentuale di identificazioni non riuscite	Identificazioni positive osservate ²	Identificazioni totali valutabili	PPC	LCL al 95% ³	UCL 95%
SNV	0	71028	0	70314	71028	98,99	98,92	99,07
Inserzione	0	1962	0	1962	1962	100	99,80	100
Delezione	0	2142	0	2098	2142	97,95	97,25	98,47

¹ Identificazione non riuscita definita come posizione cromosomica mirata in cui non è possibile determinare una variante (a causa della bassa profondità di copertura).

² Identificazione riuscita definita come posizioni cromosomiche mirate in cui viene rilevata una variante.

³ Gli intervalli di confidenza bilaterali al 95% sono calcolati utilizzando il metodo del punteggio di Wilson.

Le posizioni genomiche in cui non viene rilevata una variante somatica mirata sono riportate come negative (wild type). Per le posizioni negative previste, i dati sono stati valutati per la percentuale di identificazioni non rilevate e la percentuale di identificazioni negative. *Osservazione delle identificazioni per i risultati negativi previsti per Somatic alla pagina 75* riassume le percentuali osservate, assieme al livello superiore e inferiore di affidabilità (Lower Confidence Limit (LCL)/Upper Confidence Limit (UCL)) del 95% calcolato in base al metodo del punteggio di Wilson, per ciascun tipo di variante.

Tabella 34 Osservazione delle identificazioni per i risultati negativi previsti per Somatic

Tipo di variante	Identificazioni non riuscite osservate ¹	Identificazioni totali	Percentuale di identificazioni non riuscite	Identificazioni negative osservate ²	Identificazioni totali valutabili	PNC	LCL al 95% ³	UCL 95%
Wild type	0	92718	0	92714	92718	>99,99	99,99	100

¹ Identificazione non riuscita definita come posizione cromosomica mirata in cui non è possibile determinare una variante (a causa della bassa profondità di copertura).

² Identificazione negativa definita come posizioni cromosomiche mirate in cui non viene rilevata una variante.

³ Gli intervalli di confidenza bilaterali al 95% sono calcolati utilizzando il metodo del punteggio di Wilson.

Cronologia revisioni

Documento	Data	Descrizione della modifica
Documento n. 200025276 v01	Settembre 2022	Dati di precisione aggiornati per le osservazioni delle identificazioni della linea germinale.

Documento	Data	Descrizione della modifica
Documento n. 200025276 v00	Agosto 2022	Versione iniziale.

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, visitate 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 alla pagina Web support.illumina.com sulla scheda *Documentation* (Documentazione) per il kit in uso.