

illumina®

# TruSight Whole Genome Analysis Application

Documentazione sui prodotti

DI PROPRIETÀ DI ILLUMINA  
Documento n. 200049931 v00  
Aprile 2024  
PER USO DIAGNOSTICO IN VITRO.

## Cronologia revisioni

<b>Documento</b>	<b>Data</b>	<b>Descrizione della modifica</b>
Documento n. 200049931 v00	Aprile 2024	Versione iniziale.

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 garantire un uso sicuro e corretto dei prodotti qui descritti, le istruzioni riportate nel presente 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).

© 2024 Illumina, Inc. Tutti i diritti riservati.

Tutti i marchi di fabbrica sono di proprietà di Illumina, Inc. o dei rispettivi proprietari. Per informazioni specifiche sui marchi di fabbrica, consultare la pagina web [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).

# Sommario

Cronologia revisioni .....	ii
Descrizione generale .....	1
Informazioni preliminari .....	1
Impostazioni .....	2
Creazione della corsa .....	4
Revisione della corsa .....	5
Rimettere in coda l'analisi .....	6
Visualizzazione della corsa e dei risultati .....	7
Riepilogo file di output .....	8
Informazioni sul Report del CQ .....	8
Informazioni sul file Identificazione delle varianti .....	12
File FASTQ .....	21
File CRAM .....	22
<b>Assistenza tecnica .....</b>	<b>24</b>

# Descrizione generale

L'TruSight Whole Genome Analysis Application viene utilizzata per pianificare le corse di sequenziamento TruSight Whole Genome e avviare automaticamente l'analisi dopo il completamento della corsa. L'analisi include demultiplex, generazione di FASTQ, mappatura delle letture, allineamento al genoma di riferimento umano GrCh38/hg38 abilitato per il grafico e identificazione di varianti utilizzando Illumina DRAGEN Server for NovaSeq 6000Dx.

In diverse fasi del flusso di lavoro di analisi, l'applicazione esegue il controllo di qualità (CQ) in base alle metriche di sequenziamento, FASTQ e librerie di campioni definite e genera report con i risultati. Per i campioni che superano tutte le fasi del CQ, l'applicazione genera file di output di supporto per l'uso nelle applicazioni germline a valle.

L'TruSight Whole Genome Analysis Application esegue gli identificatori delle varianti DRAGEN, tra cui Small Variant Caller (identificatore di varianti piccole), Copy Number Variant (CNV) Caller (identificatore del numero di copie - CNV) e Repeat Expansion Detection (Rilevamento delle espansioni delle ripetizioni) con ExpansionHunter.

L'applicazione esegue anche l'annotazione del livello di affidabilità basso, intermedio o alto per le varianti piccole e include questa annotazione nel file di output.

## Informazioni preliminari

Assicurarsi che TruSight Whole Genome Analysis Application sia installata su NovaSeq 6000Dx instrument che verrà utilizzato per il sequenziamento come parte di TruSight Whole Genome. Le applicazioni installate sono disponibili nella schermata Applications (Applicazioni) su NovaSeq 6000Dx Instrument o su Illumina Run Manager utilizzando un browser su un computer in rete. Per assistenza nella pianificazione dell'installazione, contattare il rappresentante Illumina locale.

### Requisiti di archiviazione dei dati

Consultare la Documentazione del prodotto NovaSeq 6000Dx (documento n. 200010105) e la Documentazione del prodotto DRAGEN Server for NovaSeq 6000Dx (documento n. 200014171) per ulteriori informazioni sull'output e l'archiviazione dei dati.

L'TruSight Whole Genome Analysis Application genera i dati in una cartella Run (Corsa) e in una cartella Analysis (Analisi) in un archivio esterno. I requisiti minimi di archiviazione possono essere approssimati dalla dimensione dell'output dei dati in ciascuna cartella per una singola corsa di sequenziamento mostrata di seguito.

Configurazione	Cartella Run (Corsa) (GB)	Cartella Analysis (Analisi) (GB)
Cella a flusso S2 (6 campioni)	~430	~350
Cella a flusso S4 (16 campioni)	~1.110	~890

## Tempo di analisi approssimativo

L'analisi inizia automaticamente dopo il completamento di una corsa di sequenziamento e avviene in sequenza sui campioni all'interno di una corsa. I file di output dei dati saranno disponibili nella memoria esterna una volta completata l'analisi per tutti i campioni in una corsa e completato il trasferimento della copia nella memoria esterna. Quando si inizia una corsa di sequenziamento su entrambi i lati A e B contemporaneamente, il sequenziamento sarà eseguito in modo concomitante. L'analisi di queste corse di sequenziamento sarà eseguita in sequenza dall'TruSight Whole Genome Analysis Application dopo il completamento del sequenziamento. La corsa che completa prima il sequenziamento e il trasferimento sarà analizzata per prima. La seconda corsa di sequenziamento sarà trasferita e messa in coda per l'analisi dopo il completamento della prima analisi. Per informazioni su come determinare lo stato delle corse attive o non riuscite, consultare [Visualizzazione della corsa e dei risultati a pagina 7](#).

Il tempo approssimativo di disponibilità dei risultati dell'analisi dopo il completamento del sequenziamento è mostrato di seguito per la situazione in cui il lato A e il lato B vengono caricati simultaneamente con la stessa configurazione.

Configurazione	Corsa di analisi 1 (ore)	Corsa di analisi 2 (ore)
Cella a flusso S2 (6 campioni)	~12	~24
Cella a flusso S4 (16 campioni)	~24	~48

## Impostazioni

Selezionare TruSight Whole Genome Analysis Application nella schermata Applications (Applicazioni) per visualizzare la configurazione corrente e modificare le autorizzazioni.

### Configurazione

La schermata di configurazione visualizza le seguenti impostazioni dell'applicazione:

- **Nome dell'applicazione**
- **Versione dell'applicazione**
- **Versione DRAGEN**
- **Versione RTA**
- **Data di rilascio**
- **Organizzazione**
- **Identificazione del dispositivo**
- **Identificazione della produzione**

- **Library Prep Kits** (Kit di preparazione delle librerie): visualizza il kit di preparazione delle librerie. Questa impostazione non può essere modificata.
- **Index Adapter Kits** (Kit adattatori indice): visualizza i set adattatore indice disponibili per l'uso.
- **Index Reads** (Letture indici)
- **Read Type** (Tipo di lettura)
- **Index Lengths** (Lunghezze indici)
- **Read Lengths** (Lunghezze di lettura): le lunghezze di lettura sono impostate in modo predefinito quando viene selezionato l'insieme di indici. Questa impostazione non può essere modificata.

## Permessi

L'amministratore designato dispone dell'accesso alle autorizzazioni e può utilizzare le caselle di controllo nella schermata Permissions (Autorizzazioni) per gestire l'accesso degli utenti per TruSight Whole Genome Analysis Application.

Per ulteriori informazioni sulle autorizzazioni e sulla gestione degli utenti, consultare la sezione System Configuration (Configurazione del sistema) della Documentazione del prodotto NovaSeq 6000Dx (documento n. 200010105).

# Creazione della corsa

Creare nuove corse in modalità IVD sullo strumento o accedendo a Illumina Run Manager (IRM) con un browser su un computer in rete. Per accedere allo strumento da remoto, utilizzare l'indirizzo e le informazioni sull'account utente forniti dal rappresentante Illumina. Per ulteriori informazioni, consultare Documentazione del prodotto NovaSeq 6000Dx (documento n. 200010105).

Create Run (Crea corsa) è il metodo consigliato per la pianificazione della corsa. Import Sample Sheet (Importa foglio campioni) è sconsigliato. I file del foglio campioni emessi nelle cartelle di analisi e della corsa non sono adatti per l'importazione durante la pianificazione della corsa.

## Creazione di corse

1. Dalla schermata Runs (Corse), selezionare **Create Run** (Crea corsa).
2. Selezionare TruSight Whole Genome Analysis Application, quindi selezionare **Next** (Avanti).
3. Sulla schermata Run Settings (Impostazioni corsa), immettere un nome per la corsa. Il nome della corsa identifica la corsa dal sequenziamento per tutta l'analisi.

4. [Facoltativo] Immettere una descrizione per identificare la corsa. Il kit Library Prep è impostato in modo predefinito come TruSight Whole Genome e non può essere modificato.

5. Selezionare l'insieme di indici di TruSight Whole Genome desiderato dal menu a discesa **Index Adapter Kit**.

La lunghezza della lettura sarà impostata in modo predefinito e non può essere modificata. (Read [Lettura] 1 e 2 utilizzano 151 cicli; Index [Indice] 1 e 2 utilizzano 10 cicli).

6. Immettere un ID provetta della libreria (formato consigliato come DX1234567-LIB), quindi selezionare **Next** (Avanti).

Se in questa fase non è specificato alcun ID della provetta della libreria, sarà necessario selezionare la corsa pianificata prima di caricare i materiali di consumo di sequenziamento. Se in questa fase viene immesso un ID della provetta della libreria errato, la corsa pianificata deve essere corretta prima di caricare i materiali di consumo. Consultare [Revisione della corsa a pagina 5](#) per il protocollo per correggere la corsa quando si è pronti a caricare i materiali di consumo.

7. Nelle schermate Sample Data (Dati campione) e Sample Settings (Impostazioni campione), vengono immesse le informazioni sul campione. I dati del campione possono essere inseriti manualmente o importando un file di campioni. L'ID campione deve essere univoco per ciascun campione e può contenere solo caratteri alfanumerici, trattini bassi e trattini. Non includere spazi. Well Position (Posizione pozzetto) si riferisce al pozzetto nel formato da A01 a H04 della piastra indici. Le informazioni sulla sequenza degli indici saranno compilate automaticamente quando viene inserita Well Position (Posizione del pozzetto) della piastra indici. Il sesso deve essere inserito come Male (Maschio) Female (Femmina) o Unknown (Sconosciuto). Library Plate ID (ID piastra libreria) e Library Well ID (ID pozzetto libreria) (ad es, formato A01) sono campi obbligatori.



- Per inserire i dati del campione manualmente, aggiungere delle righe (per un totale di 6 per la cella a flusso S2 o 16 per la cella a flusso S4) e inserire le informazioni richieste nei campi Sample ID (ID campione) e Well Position (Posizione pozzetto). Le informazioni possono anche essere copiate e incollate da Excel. Selezionare **Next** (Avanti). Nella schermata Sample Settings (Impostazioni campione), immettere Library Plate ID (ID piastra libreria), Library Well ID (ID pozzetto libreria) e Sex (Sesso). Selezionare **Next** (Avanti).
  - Per importare un file di dati campione, selezionare **Import Samples** (Importa campioni) e caricare il file di dati campione. Le informazioni verranno inserite automaticamente nelle righe. In questa schermata è disponibile un modello (\*.csv). Selezionare **Next** (Avanti). Nella schermata Sample Settings (Impostazioni campione), le informazioni verranno inserite automaticamente in righe dal file di dati campione importato. Selezionare **Next** (Avanti).
8. Nella schermata Analysis Settings (Impostazioni analisi), immettere il nome del lotto registrato durante la pianificazione del lotto e della corsa.
  9. [Facoltativo] Selezionare Flow Cell Type (Tipo di cella a flusso), S2 o S4.
  10. Confermare o deselezionare la casella di controllo per generare FASTQ con compressione ORA, quindi selezionare **Next** (Avanti).

**NOTA** TruSight Whole Genome Analysis Application genera FASTQ con compressione ORA per impostazione predefinita. La modifica di questa impostazione aumenterà le dimensioni dell'output finale dei dati.

11. Nella schermata Run Review (Revisione corsa), controllare le informazioni immesse. Se non sono necessarie modifiche, selezionare **Save** (Salva). Se sono necessarie modifiche, selezionare **Back** (Indietro) per tornare alla schermata appropriata.



### ATTENZIONE

TruSight Whole Genome è stato convalidato per 6 campioni quando si utilizza la cella a flusso NovaSeq 6000Dx S2 e 16 campioni quando si utilizza la cella a flusso NovaSeq 6000Dx S4. Assicurarsi di immettere il numero corretto di campioni per la configurazione della cella a flusso selezionata.

## Revisione della corsa

Se sono necessarie modifiche dopo la creazione della corsa e prima di caricare i materiali di consumo per il sequenziamento, controllare le analisi in modalità IVD sullo strumento o accedendo a Illumina Run Manager (IRM) utilizzando un browser su un computer in rete.

1. Selezionare **Runs** (Corse).
2. Selezionare il nome della corsa nella scheda Planned Runs (Corse pianificate).
3. Selezionare **Edit** (Modifica).

4. Aggiornare le informazioni sulla corsa o sul campione secondo necessità. Ad esempio, immettere o correggere l'ID della provetta della libreria in modo che corrisponda a quello utilizzato durante il completamento del flusso di lavoro.
5. Selezionare **Next** (Avanti) fino a Run Review (Revisione corsa).
6. Selezionare **Save** (Salva).
7. Selezionare **Exit** (Esci).

Tornare al sequenziamento in modalità IVD per ripetere il caricamento dei materiali di consumo. La corsa dovrebbe ora essere evidenziata automaticamente.

Se si aggiorna l'ID della provetta della libreria durante il caricamento dei materiali di consumo, tornare a Run Selection (Selezione corsa) nel software di controllo e selezionare **Refresh** (Aggiorna) per la colonna associata, A o B. La corsa dovrebbe ora essere evidenziata automaticamente. In caso contrario, selezionare **Back** (Indietro) per ripetere il caricamento dei materiali di consumo.

## Rimettere in coda l'analisi

Consultare la sezione Risoluzione dei problemi nella Foglio illustrativo di TruSight Whole Genome (documento n. 200050132) per determinare il tipo di rimessa in coda dell'analisi più appropriato.

### Rimettere in coda l'analisi senza modifiche

1. Selezionare Completed Run (Corsa completata) per visualizzare Run Details (Dettagli corsa).
2. Selezionare **Requeue Analysis** (Rimetti in coda l'analisi).
3. Selezionare **Requeue Analysis with no changes** (Rimetti in coda l'analisi senza modifiche).
4. Fornire i dettagli nel campo Reanalysis Reason (Motivo della ripetizione dell'analisi).
5. Selezionare **Requeue Analysis** (Rimetti in coda l'analisi).
6. Uscire dalla pagina e andare alla pagina Active Runs (Corse attive) per confermare che la rimessa in coda è in corso.

### Rimettere in coda l'analisi senza modifiche

1. Selezionare Completed Run (Corsa completata) per visualizzare Run Details (Dettagli corsa).
2. Selezionare **Requeue Analysis** (Rimetti in coda l'analisi).
3. Selezionare **Edit run settings** (Modifica impostazioni corsa) e **Requeue Analysis** (Rimetti in coda l'analisi).
4. Fornire i dettagli nel campo Reanalysis Reason (Motivo della ripetizione dell'analisi).
5. Selezionare **Requeue Analysis** (Rimetti in coda l'analisi).
6. Confermare o aggiornare Run Settings (Impostazioni della corsa), quindi selezionare **Next** (Avanti).

7. Correggere le informazioni del campione secondo necessità aggiornando manualmente i campi o selezionare **Download Template** (Scarica modello) per creare un file `sampledata.csv` con le informazioni attuali. Correggere le informazioni ed eliminare le righe esistenti nella scheda Sample Data (Dati campione) prima di utilizzare Import Samples (Importa campioni) per compilare i dati campione corretti.
8. Rivedere le informazioni nella schermata Run Review (Revisione corsa) e selezionare **Save** (Salva) per iniziare la ripetizione dell'analisi.
9. Selezionare **Exit** (Esci) e andare alla pagina Active Runs (Corse attive) per confermare che è in corso la rimessa in coda.

La cartella dei dati di analisi originale deve essere presente nella posizione di archiviazione esterna specificata in Run Details (Dettagli corsa) per essere completata correttamente. Se la ripetizione dell'analisi non va a buon fine, assicurarsi che la corsa non sia stata spostata o eliminata.

## Visualizzazione della corsa e dei risultati

1. Dalla schermata principale di Illumina Run Manager in modalità IVD, selezionare **Runs** (Corse).
2. Dalla scheda Completed Runs (Corse completate), selezionare il nome della corsa.  
Questa scheda visualizzerà anche le corse completate a causa di un errore di sequenziamento, trasferimento o analisi dei dati. Le corse attive e il relativo stato vengono visualizzati nella scheda Active Runs (Corse attive). Per ulteriori informazioni, consultare Documentazione del prodotto NovaSeq 6000Dx (documento n. 200010105).
3. Selezionare il nome della corsa nella scheda Completed Runs (Corse completate) per visualizzare i dettagli e i risultati della corsa per il percorso della cartella Analysis Output (Output dell'analisi).  
Per le corse non andate a buon fine, controllare lo stato per ogni fase e consultare la sezione Risoluzione dei problemi del Foglio illustrativo di TruSight Whole Genome (documento n. 200050132).
4. Andare alla cartella di analisi sull'unità locale e aprire Consolidated Report (Report consolidato) per controllare il risultato SUPERATO/NON SUPERATO per ogni fase del CQ come segue:
  - Per il CQ della corsa di sequenziamento, consultare Riepilogo del risultato del CQ del sequenziamento
  - Per il CQ FASTQ per ciascun campione della serie, consultare Risultato del riepilogo del CQ FASTQ
  - Per il CQ delle librerie per ciascun campione nell'analisi, consultare Risultato del CQ del riepilogo delle librerie di campioniSe si osserva un risultato NON SUPERATO, annotare la fase del CQ e consultare la sezione Risoluzione dei problemi del Foglio illustrativo di TruSight Whole Genome (documento n. 200050132).

## Riepilogo file di output

TruSight Whole Genome Analysis Application salva i seguenti file di output principali. Per la posizione dei file di output principali, consultare le sezioni relative alle informazioni sui file riportate di seguito.

Le analisi e i campioni che non superano i criteri di validità non producono file CRAM, ROH bed o \*genome.vcf.

File di output	Descrizione
Report consolidato (*.csv)	Contiene le metriche di qualità utilizzate per la validità della corsa (inclusi resa totale e Q30), le metriche di validità dei campioni (inclusa resa di FASTQ), le metriche CQ delle librerie e le metriche Solo per informazioni (FIO) per tutti i campioni dell'analisi.
Report campioni (*.csv)	Contiene i risultati del CQ del sequenziamento, FASTQ e del CQ della librerie di campioni. Il report contiene anche la concordanza della ploidia e le metriche FIO per il singolo campione, nonché la corsa di sequenziamento associata.
VCF delle varianti piccole e mSNV (*.annotated.hard-filtered-gvcf.gz)	Contiene informazioni di identificazione sulle varianti per le varianti piccole (SNV, indel) e le SNV mitocondriali.
VCF del CNV (*.cnv.vcf.gz)	Contiene informazioni sull'identificazione delle varianti per guadagni e perdite del numero di copie
Ripetizioni del VCF (*.repeats.vcf.gz)	Contiene informazioni sulle identificazioni delle varianti sulle espansioni delle STR e SMN1.
ROH BED (*.roh.bed)	Contiene informazioni per le regioni di omozigosi.
FASTQ (*.fastq.gz o *.fastq.ora)	File intermedi che contengono le identificazioni delle basi qualitativamente valutate. I file FASTQ rappresentano gli input principali per la fase di allineamento. Se è stata selezionata la compressione ORA, il nome del file riflette questa scelta.
CRAM allineamento (*.cram)	Contiene le letture allineate per un determinato campione.

## Informazioni sul Report del CQ

Il Consolidated Report (Report consolidato) <<RunID>>\_Consolidated\_Report.csv si trova nella directory TruSightWholeGenomeAnalysis\_x.x.x\_run-complete e contiene informazioni sulle metriche di qualità utilizzate per valutare il superamento o il mancato superamento dei campioni in

diverse fasi dell'analisi. I singoli report dei campioni <<Sample\_ID>>\_Sample\_Report.csv sono disponibili nelle cartelle <Sample\_ID> nella directory TruSightWholeGenomeAnalysis\_x.x.x\_run-complete.

Le intestazioni del report includono le seguenti informazioni sulla corsa: versione dell'app, nome del lotto, ID provetta del pool di librerie, nome della corsa di sequenziamento, ID corsa di sequenziamento e tipo di cella a flusso. Le tabelle seguenti descrivono le informazioni incluse nel Consolidated Report (Report consolidato). Il singolo Sample Report (Report campioni) include le stesse informazioni, ad eccezione delle metriche Demultiplex.

Tabella 1 Metriche del controllo qualità del sequenziamento

Metriche	Spec.	Descrizione
Resa totale non indicizzata (GB)	NA	Nessuna specifica poiché le corse a resa inferiore possono risultare nel superamento delle librerie di campioni. Prevedere $\geq 3.000$ Gbp per la cella a flusso S4 e $\geq 1.000$ GB per la cella a flusso S2.
% totale $\geq$ Q30	$\geq 85$	Misura della qualità delle basi a livello della corsa. Viene impostata una specifica minima poiché cicli %Q30 troppo bassi non supereranno le basi Q30 nel CQ delle librerie di campioni.
Riepilogo del risultato del controllo qualità del sequenziamento	SUPERATO o NON SUPERATO	Per il mancato superamento del controllo qualità del sequenziamento, consultare la sezione Risoluzione dei problemi nella Foglio illustrativo di TruSight Whole Genome (documento n. 200050132).

Tabella 2 Metriche demultiplex

Metriche	Spec.	Descrizione
Percentuale di letture identificate	NA	Frazione totale di letture del superamento del filtro nella corsa assegnate ai campioni durante il demultiplex.
CV percentuale	NA	Fornisce una misura dell'uniformità delle letture demultiplex a ciascuna coppia di indici sulla corsa. Prevedere $< 25\%$ per le corse senza errori dei risultati CQ FASTQ.

Tabella 3 Metriche del CQ FASTQ

<b>Metriche</b>	<b>Spec.</b>	<b>Descrizione</b>
Resa per campione (bps)	$\geq 90.000.000.000$	Il valore minimo è impostato per essere equivalente a una copertura autosomica media ~26x al fine di eseguire lo smistamento delle librerie di campioni che non supereranno il CQ per ridurre il tempo di analisi.
Risultato del riepilogo del CQ FASTQ	SUPERATO o NON SUPERATO	Per il mancato superamento del CQ FASTQ, consultare la sezione Risoluzione dei problemi nella Foglio illustrativo di TruSight Whole Genome (documento n. 200050132).

Tabella 4 Metriche del CQ qualità delle librerie di campioni

<b>Metriche</b>	<b>Spec.</b>	<b>Descrizione</b>
Copertura autosomica media	$\geq 35$	Copertura media tra gli autosomi. La specifica minima è impostata per garantire le prestazioni analitiche.
Percentuale di autosomi con copertura superiore a 20X	$\geq 93,94$	Misura dell'uniformità di copertura che rileva problemi non necessariamente correlati alla distorsione GC. La specifica minima è impostata per garantire le prestazioni analitiche.
Copertura normalizzata dal 60% al 79% degli intervalli GC	$0,82 \leq x \leq 1,13$	Misura dell'uniformità di copertura che rileva la distorsione GC, in particolare una perdita di copertura in aree del genoma con una % di GC più alta e una % di composizione delle basi AT inferiore. Le specifiche minime e massime sono impostate per garantire le prestazioni analitiche.
Copertura normalizzata dal 20% al 39% degli intervalli GC	$0,97 \leq x \leq 1,06$	Misura dell'uniformità di copertura che rileva la distorsione GC, in particolare una perdita di copertura in aree del genoma con una % di GC inferiore e una % di composizione di basi AT superiore. Le specifiche minime e massime sono impostate per garantire le prestazioni analitiche.
Copertura mitocondriale media	$\geq 500$	Copertura del cromosoma mitocondriale. La specifica minima è impostata per garantire il limite di rilevamento delle SNV mitocondriali.
Percentuale basi Q30	$\geq 85$	Misura della qualità delle basi. La specifica minima è impostata per garantire le prestazioni analitiche.

<b>Metriche</b>	<b>Spec.</b>	<b>Descrizione</b>
Contaminazione stimata del campione	≤ 0,005	Rileva letture contaminanti da altri campioni. La specifica massima è impostata per garantire il limite di rilevamento delle SNV mitocondriali (il tipo di variante con la massima sensibilità alla contaminazione).
Risultato del CQ del riepilogo delle librerie di campioni	SUPERATO o NON SUPERATO	Per il mancato superamento del CQ delle librerie di campioni, consultare la sezione Risoluzione dei problemi nella Foglio illustrativo di TruSight Whole Genome (documento n. 200050132).

Tabella 5 Metriche del CQ della ploidia

<b>Metriche</b>	<b>Spec.</b>	<b>Descrizione</b>
Ploidia dei cromosomi sessuali fornita	NA	Sesso fornito dall'operatore durante la creazione della corsa (Female [Femmina], Male [Maschio], Unknown [Sconosciuto]).
Stima della ploidia	NA	Ploidia sessuale stimata da DRAGEN.
Riepilogo dei risultati della ploidia	CONCORDANT (CONFORME), DISCORDANT (NON CONFORME) o ND	CONCORDANT (CONFORME) indica la concordanza tra la ploidia sessuale fornita e stimata. ND indica il sesso fornito come Unknown (Sconosciuto) o una stima diversa da XX o XY. Per i risultati DISCORDANT (NON CONFORME), è stato inserito un sesso errato durante la creazione dell'analisi o potrebbe essersi verificato uno scambio di campioni. Consultare la sezione Risoluzione dei problemi nel Foglio illustrativo di TruSight Whole Genome (documento n. 200050132).

Tabella 6 Per le metriche Solo informazioni

<b>Metriche</b>	<b>Descrizione</b>
Lunghezza mediana dell'inserzione	Il target è di 450 bp, ma si prevede una variazione per corsa di sequenziamento e operatore. È accettabile un intervallo compreso tra circa 360 e 550 bp. Un funzionamento costante al di fuori di questo intervallo può portare a una maggiore incidenza di mancato superamento del campione.
Percentuale di letture mappate	Percentuale di letture che si mappano sul genoma di riferimento. Può essere ridotta in risposta a contaminazione di gDNA non umano, scarsa qualità del campione o lunghezza dell'inserzione troppo piccola che causa problemi di mappatura.

<b>Metriche</b>	<b>Descrizione</b>
Percentuale di letture con allineamenti supplementari	Percentuale di letture con mappatura che si suddivide in diverse posizioni nel genoma di riferimento.
Percentuale di letture contrassegnate duplicate	Prevedere < 20%. Può essere elevata se la resa di preparazione delle librerie è bassa o il raggruppamento in pool è inferiore al volume richiesto o in risposta a problemi correlati al sequenziamento.
Percentuale basi soft clipped Read 1 (Lettura 1)	Utile nella diagnosi della causa principale del mancato superamento della copertura autosomica media.
Percentuale basi soft clipped Read 2 (Lettura 2)	Utile nella diagnosi della causa principale del mancato superamento della copertura autosomica media.
Percentuale basi tagliate Read 1 (Lettura 1)	Utile nella diagnosi della causa principale del mancato superamento della copertura autosomica media.
Percentuale basi tagliate Read 2 (Lettura 2)	Utile nella diagnosi della causa principale del mancato superamento della copertura autosomica media.

## Informazioni sul file Identificazione delle varianti

### File VCF

I file Variant call format (\*.vcf) contengono informazioni sulle varianti trovate in posizioni specifiche in un genoma di riferimento e sono disponibili nella directory <Sample\_ID>/Analysis.

L'intestazione del file VCF include la versione del formato del file VCF e la versione dell'identificatore di varianti ed elenca le annotazioni utilizzate nel resto del file. L'ultima riga dell'intestazione contiene le intestazioni delle colonne per le righe dei dati. Ogni riga di dati del file VCF contiene informazioni su una singola posizione di riferimento.

Tutti i file VCF contengono un'intestazione con le descrizioni delle colonne di output e i dati delle identificazioni delle varianti nelle colonne etichettate come CHROM, POS, ID, REF, ALT, QUAL, FILTRO, INFO, FORMATO, CAMPIONE. Le definizioni dei valori delle colonne possono variare a seconda degli identificatori delle varianti.



## Variante piccola e mSNV VCF

L'output viene salvato nel file `<Sample_ID>.annotated.hard-filtered.gvcf.gz` nella directory `<Sample_ID>/Analysis`.

Un file VCF genomico (genomic VCF, gVCF) contiene informazioni su varianti e posizioni ritenute omozigoti rispetto al genoma di riferimento. Per le regioni omozigoti, il file gVCF include statistiche che indicano in che misura le letture supportano l'assenza di varianti o alleli alternativi. Il file gVCF include un allele `<NON_REF>` artificiale. Alle letture che non supportano il riferimento o qualsiasi variante viene assegnato l'allele `<NON_REF>`. DRAGEN utilizza queste letture per determinare se la posizione può essere identificata come riferimento omozigote, invece di rimanere non identificata. Il punteggio risultante rappresenta il livello di affidabilità su scala Phred in un'identificazione di riferimento omozigote. In modalità germline, il punteggio è FORMAT/GQ.

DRAGEN fornisce il filtraggio delle varianti post-VCF in base alle annotazioni presenti nei record VCF. Il filtraggio rigido delle varianti è descritto di seguito. Tuttavia, a causa della natura degli algoritmi di DRAGEN, che incorporano l'ipotesi di errori correlati dall'interno del nucleo dell'identificatore delle varianti, la pipeline ha migliorato le capacità nel distinguere le varianti reali dal rumore, e pertanto la dipendenza dal filtraggio post-VCF è sostanzialmente ridotta.

TruSight Whole Genome Analysis Application fornisce l'annotazione del punteggio di affidabilità e del livello di affidabilità per le varianti piccole che possono essere utilizzate per migliorare ulteriormente le prestazioni. L'annotazione del livello di affidabilità non è un filtro di qualità e pertanto non si riflette direttamente nello stato di qualità delle identificazioni delle varianti. Pertanto, è possibile vedere le identificazioni delle varianti che superano il filtro che sono comunque annotate come di scarsa affidabilità.

Tabella 7 Intestazioni dei file VCF

Intestazione	Descrizione
CHROM	Il cromosoma del genoma di riferimento. I cromosomi appaiono nello stesso ordine del file FASTA di riferimento.
POS	La posizione a singola base della variante nel cromosoma di riferimento. Per le varianti a singolo nucleotide (SNV), questa posizione è la base di riferimento con la variante. Per le indel, questa posizione è la base di riferimento immediatamente precedente la variante.
ID	Sempre .
REF	Il genotipo di riferimento. Ad esempio, una delezione di una singola T è rappresentata come TT di riferimento e T alternativa. Una variante da A a T a singolo nucleotide è rappresentata come A di riferimento e T alternativa.

Intestazione	Descrizione
ALT	Gli alleli che differiscono dalla lettura di riferimento. Ad esempio, l'inserimento di una singola T viene rappresentato come A di riferimento e AT alternativo. Una variante da A a T a singolo nucleotide è rappresentata come A di riferimento e T alternativa.
QUAL	Un punteggio di qualità con scala Phred assegnato dall'identificatore della variante. Punteggi più alti indicano una maggiore affidabilità nella variante e una minore probabilità di errori. Per un punteggio qualitativo di Q, la probabilità di errore stimata è $10^{-(Q/10)}$ . Ad esempio, l'insieme delle identificazioni Q30 ha un tasso di errore dello 0,1%. Molti identificatori di varianti assegnano punteggi di qualità basati sui loro modelli statistici, che sono elevati in relazione al tasso di errore osservato.

Tabella 8 Annotazioni dei file VCF

Intestazione	Descrizione
FILTRO	<ul style="list-style-type: none"> <li>• SUPERATO: Tutti i filtri sono superati.</li> <li>• <b>DRAGENSnpHardQUAL</b>: impostare se true:QUAL &lt; 10,41</li> <li>• <b>DRAGENIndelHardQUAL</b>: impostare se true:QUAL &lt; 7,83</li> <li>• <b>LowDepth</b>: impostare se true:DP &lt;= 1</li> <li>• <b>LowGQ</b>: impostare se true:GQ = 0</li> <li>• <b>PloidyConflict</b>: identificazione di genotipo dall'identificatore di varianti non coerente con la ploidia cromosomica</li> <li>• <b>base_quality</b>: sito filtrato perché la qualità mediana delle basi delle letture alt in questo locus non soddisfa la soglia</li> <li>• <b>filtered_reads</b>: sito filtrato perché è stata filtrata una frazione troppo grande di letture</li> <li>• <b>fragment_length</b>: sito filtrato perché la differenza assoluta tra la lunghezza mediana dei frammenti delle letture alt e la lunghezza mediana dei frammenti delle letture ref a questo locus supera la soglia</li> <li>• <b>low_af</b>: la frequenza allelica non soddisfa la soglia</li> <li>• <b>low_depth</b>: sito filtrato perché la profondità di lettura è troppo bassa</li> <li>• <b>low_frac_info_reads</b>: sito filtrato perché la frazione di letture informative è inferiore alla soglia</li> <li>• <b>low_normal_depth</b>: sito filtrato perché la profondità di lettura normale del campione è troppo bassa</li> <li>• <b>long_indel</b>: sito filtrato perché la lunghezza dell'indel è troppo lunga</li> <li>• <b>mapping_quality</b>: sito filtrato perché la qualità di mappatura mediana delle letture alt a questo locus non soddisfa la soglia</li> <li>• <b>multiallelic</b>: sito filtrato perché più di due alleli alt superano il LOD del tumore</li> <li>• <b>non_homref_normal</b>: sito filtrato perché il genotipo del campione normale non è omozigote di riferimento</li> <li>• <b>no_reliable_supporting_read</b>: sito filtrato perché non esiste una lettura somatica di supporto affidabile</li> <li>• <b>panel_of_normals</b>: osservato in almeno un campione del pannello delle normalità vcf</li> <li>• <b>read_position</b>: sito filtrato perché la mediana delle distanze tra l'inizio e la fine della lettura e questo locus è inferiore alla soglia</li> <li>• <b>RMxNRepeatRegion</b>: sito filtrato perché tutto o parte dell'allele della variante è una ripetizione del riferimento</li> <li>• <b>strand_artifact</b>: sito filtrato a causa di un grave bias del filamento</li> <li>• <b>str_contraction</b>: sito filtrato a causa di un sospetto errore di PCR in cui l'allele alt è inferiore di un'unità di ripetizione rispetto al riferimento</li> <li>• <b>too_few_supporting_reads</b>: sito filtrato perché ci sono troppo poche letture di supporto nel campione di tumore</li> <li>• <b>weak_evidence</b>: il punteggio della variante somatica non soddisfa la soglia</li> </ul>

Intestazione	Descrizione
INFO	<ul style="list-style-type: none"> <li>• <b>DB</b>: appartenenza a dbSNP.</li> <li>• <b>FS</b>: valore p scalato con Phred mediante il test esatto di Fisher per rilevare il bias del filamento.</li> <li>• <b>QD</b>: l'affidabilità/qualità della variante per la profondità.</li> <li>• <b>R2_5P_bias</b>: punteggio basato sul bias di accoppiamento e sulla distanza dall'estremità principale 5.</li> <li>• <b>SOR</b>: rapporto di probabilità simmetrico della tabella di contingenza 2x2 per rilevare il bias del filamento.</li> <li>• <b>DP</b>: profondità di lettura approssimativa (informativa e non informativa); alcune letture possono essere state filtrate in base a mapq ecc.</li> <li>• <b>END</b>: posizione di arresto dell'intervallo.</li> <li>• <b>FractionInformativeReads</b>: la frazione di letture informative sul totale delle letture.</li> <li>• <b>MQ</b>: qualità della mappatura RMS.</li> <li>• <b>MQRankSum</b>: punteggio Z dal test della somma dei ranghi di Wilcoxon delle qualità di mappatura delle letture Alt rispetto a Ref.</li> <li>• <b>ReadPosRankSum</b>: punteggio Z del test di somma dei ranghi di Wilcoxon sulla polarizzazione delle posizioni di lettura Alt rispetto a Ref.</li> <li>• <b>SOMATIC</b>: almeno una variante in questa posizione è somatica.</li> <li>• <b>IENS</b>: lunghezze Indel per ciascuna variante ALT.</li> <li>• <b>SCORE</b>: punteggio di affidabilità per ciascun tipo di variante presente presso il centro come (tipo di variante):(punteggio di affidabilità).</li> <li>• <b>TIER</b>: livello di affidabilità per ciascun tipo di variante presente presso il centro come (tipo di variante):(livello di affidabilità).</li> </ul>

Intestazione	Descrizione
FORMATO	<p>La colonna FORMAT (FORMATO) elenca i campi separati da due punti.</p> <ul style="list-style-type: none"> <li>• <b>AD</b>: profondità alleliche (contando solo le letture informative sul totale delle letture) per gli alleli ref e alt nell'ordine elencato.</li> <li>• <b>AF</b>: frazioni alleliche per gli alleli alt nell'ordine elencato.</li> <li>• <b>DP</b>: la profondità approssimativa della lettura (letture con MQ=255 o con accoppiamenti non corretti sono filtrate).</li> <li>• <b>F1R2</b>: conteggio delle letture in orientamento delle coppie F1R2 che supportano ciascun allele.</li> <li>• <b>F2R1</b>: conteggio delle letture in orientamento delle coppie F2R1 che supportano ciascun allele.</li> <li>• <b>GP</b>: probabilità posteriori scalate con Phred per i genotipi, come definito nella specifica del VCF.</li> <li>• <b>GQ</b>: qualità del genotipo.</li> <li>• <b>GT</b>: il genotipo.</li> <li>• <b>ICNT</b>: conteggi delle letture informative INDEL in base al modello di affidabilità di riferimento.</li> <li>• <b>MB</b>: statistiche delle componenti per campione per rilevare i bias di accoppiamento.</li> <li>• <b>MIN_DP</b>: DP minimo osservato all'interno del blocco GVCF.</li> <li>• <b>PL</b>: probabilità normalizzate e scalate con Phred per i genotipi, come definito nelle specifiche del VCF.</li> <li>• <b>PRI</b>: probabilità antecedenti scalate con Phred per i genotipi.</li> <li>• <b>PS</b>: informazioni sull'ID fisico di phasing, dove ogni ID univoco all'interno di un dato campione (ma non tra i campioni) collega i record all'interno di un gruppo di phasing.</li> <li>• <b>SB</b>: statistiche dei componenti per campione che comprendono il test esatto di Fisher per rilevare il bias del filamento.</li> <li>• <b>SPL</b>: probabilità normalizzate, scalate con Phred per gli SNP in base al modello di affidabilità di riferimento.</li> <li>• <b>SQ</b>: qualità di Somatic.</li> </ul>
CAMPIONE	La colonna del campione fornisce i valori specificati nella colonna FORMAT (FORMATO).

## VCF delle varianti del numero di copie

Lo stadio dei conteggi dei target è il primo stadio di elaborazione per la pipeline DRAGEN CNV, che produce `<Sample_ID>.target.counts.gz`, quindi viene eseguita la correzione della distorsione GC, generando un file `*.target.counts.gc-corrected.gz`. La fase di normalizzazione produce il file `*.tn.tsv.gz`. Il software DRAGEN Host genera molti file intermedi. `*.seg.called.merged` è il file di identificazione finale che contiene gli eventi di amplificazione e delezione.

Oltre al file del segmento, DRAGEN emette le identificazioni nel formato VCF standard. L'output viene salvato in `<Sample_ID>.cnv.vcf.gz` nella directory `<Sample_ID>/Analysis`.

Definizioni di colonne specifiche per l'identificatore delle CNV:

La colonna POS è la posizione iniziale della variante. Secondo la specifica del VCF, se uno qualsiasi degli alleli ALT è un allele simbolico, come `<DEL>`, allora è necessaria la base padding e POS indica la coordinata della base che precede il polimorfismo. Tutte le coordinate nel VCF sono basate su 1.

La colonna ID viene utilizzata per rappresentare l'evento. Il campo ID codifica il tipo di evento e le coordinate dell'evento.

La colonna REF contiene una N per tutti gli eventi CNV.

La colonna ALT specifica il tipo di evento CNV. Poiché il VCF contiene solo eventi CNV, viene utilizzata solo la voce DEL o DUP.

La colonna QUAL contiene un punteggio di qualità stimato per l'evento CNV, utilizzato nel filtraggio rigido.

La colonna FILTER (FILTRO) contiene PASS (SUPERATO) se l'evento CNV supera tutti i filtri, altrimenti la colonna contiene il nome del filtro non superato.

La colonna INFO contiene informazioni che rappresentano l'evento. La voce REFLEN indica la durata dell'evento. La voce SVTYPE è sempre CNV. La voce END (FINE) indica la posizione finale dell'evento.

I campi FORMAT (FORMATO) sono descritti nell'intestazione.

- GT: il genotipo.
- SM: rapporto delle copie lineare della media del segmento
- CN: numero di copie stimato
- BC: numero di intervalli nella regione
- PE: numero di letture finali non correttamente appaiate ai breakpoint di avvio e arresto

## Ripetizioni VCF

ExpansionHunter esegue un riallineamento basato su grafico sequenziale delle letture che hanno origine all'interno e intorno a ciascuna ripetizione target. ExpansionHunter quindi genotipizza la lunghezza della ripetizione in ciascun allele in base a questi allineamenti grafici.

Ulteriori informazioni e analisi sono disponibili nei seguenti documenti di ExpansionHunter:

- Dolzhenko et al. *Detection of long repeat expansions from PCR-free whole-genome sequence data* 2017
- Dolzhenko et al. *ExpansionHunter: A sequence-graph based tool to analyze variation in short tandem repeat regions* 2019

Il catalogo delle varianti TruSight Whole Genome Analysis Application STR contiene specifiche sulle ripetizioni che causano le malattie situate nei geni AFF2, AR, ATN1, ATXN1, ATXN10, ATXN2, ATXN3, ATXN7, ATXN8OS, C9ORF72, CACNA1A, CBL, CNBP, CSTB, DIP2B, DMPK, FMR1, FXN, GLS, HTT, JPH3, NIPA1, NOP56, NOTCH2NL, PABPN1, PHOX2B, PPP2R2B e TBP.

I risultati della genotipizzazione ripetuta vengono emessi come file VCF separato, che fornisce la lunghezza di ciascun allele ad ogni ripetizione identificabile definita nel file di catalogo delle specifiche della ripetizione. Il nome del file è <Sample\_ID>.repeats.gz e si trova nella directory <Sample\_ID>/Analysis.

Alcune colonne sono specifiche per la ripetizione dell'identificatore delle espansioni delle ripetizioni:

Tabella 9 Campi VCF principali

Campo	Descrizione
CHROM	Identificatore dei cromosomi
POS	Posizione della prima base prima della regione di ripetizione nel riferimento
ID	Sempre .
REF	La base di riferimento nella posizione POS
ALT	Elenco degli alleli ripetuti nel formato <STR <sub>n</sub> >. N è il numero di unità ripetute.
QUAL	Sempre .
FILTRO	Il filtro LowDepth viene applicato quando la profondità complessiva del locus è inferiore a 10x o il numero di letture che si estendono su uno o entrambi i breakend è inferiore a 5.

Tabella 10 Campi INFO aggiuntivi

Campo	Descrizione
END	Posizione dell'ultima base della regione di ripetizione nel riferimento
REF	Numero di copie di riferimento
REPID	ID variante dal catalogo delle varianti
RL	Lunghezza di riferimento in bp
RU	Unità di ripetizione con l'orientamento di riferimento
VARID	ID variante dal catalogo delle varianti

Tabella 11 Campi GENOTYPE (per campione)

Campo	Descrizione
AD	Profondità alleliche per gli alleli di riferimento e alt nell'ordine elencato
ADFL	Numero di letture di affiancamento coerenti con l'allele

Campo	Descrizione
ADIR	Numero di letture in ripetizione coerenti con l'allele
ADSP	Numero di letture spanning coerenti con l'allele
DST	Risultati (+ rilevati, - non rilevati, ? indeterminati) del test rappresentati dalla variante
GT	Genotipo
LC	Copertura dei loci
REPCI	Intervallo di confidenza per REPCN
REPCN	Numero di unità ripetute coperte dall'allele
RPL	Report di probabilità logaritmica per la presenza dell'allele di riferimento
SO	Tipo di letture che supportano l'allele. I valori possono essere SPANNING, FLANKING o INREPEAT. Questi valori indicano se le letture si estendono (spanning), si affiancano (flanking) o sono completamente contenute nella ripetizione (in repeat).

Il file <Sample\_ID>.repeats.vcf.gz include l'output SMN insieme a eventuali ripetizioni mirate. L'output SMN è rappresentato come una singola identificazione delle SNV nella posizione che influisce sulla scissione in SMN1 (NM\_000344.3:c.840C/T) con stato di atrofia muscolare spinale (SMA) nei seguenti campi personalizzati.

Tabella 12 Risultati SMA nel file di output repeats.vcf

Campo	Descrizione
VARID	SMN contrassegna l'identificazione dell'SMN.
GT	Identificazione del genotipo in questa posizione utilizzando un modello di genotipo normale (diploide).
DST	Identificazione dello stato SMA: + indica rilevato - indica non rilevato ? indica indeterminato
AD	Conteggi di lettura totali che supportano gli alleli C e T.
RPL	Report di probabilità Log10 tra i modelli non interessati e quelli interessati. I punteggi positivi indicano che il modello non interessato è più probabile.

## ROH BED

Le regioni di omozigosi (ROH) vengono rilevate come parte dell'identificatore della variante piccola. L'identificatore rileva e produce le corse di omozigosi dalle identificazioni dell'intero genoma sui cromosomi umani autosomici. I cromosomi sessuali vengono ignorati a meno che il cariotipo sessuale campione non sia XX, come determinato dal Ploidy Estimator (Strumento di stima della ploidia). L'output



delle ROH consente agli strumenti a valle di effettuare lo screening e predire la consanguineità tra i genitori del soggetto probando.

Una regione è definita come identificazioni di una variante consecutiva sul cromosoma senza un ampio spazio tra queste varianti. In sintesi, le regioni sono spezzate dal cromosoma o da grandi lacune vuote senza identificazioni delle SNV. La dimensione delle lacune è impostata su 3 Mbase.

L'identificatore delle ROH produce un file di output ROH denominato `<Sample_ID>.roh.bed` nella directory `<Sample_ID>/Analysis`. Ogni riga rappresenta una regione di omozigosi. Il file bed contiene le seguenti colonne:

```
Chromosome Start End Score #Homozygous #Heterozygous
```

Dove

- Score (Punteggio) è una funzione del numero di varianti omozigoti ed eterozigoti, in cui ciascuna variante omozigote aumenta il punteggio di 0,025 e ciascuna variante eterozigote riduce il punteggio di 0,975.
- Le posizioni Start (Iniziale) ed End (Finale) sono un intervallo di semiapertura basato su 0.
- #Homozygous (N. omozigosi) è il numero di varianti omozigoti nella regione.
- #Heterozygous (N. eterozigosi) è il numero di varianti eterozigoti nella regione.

L'identificatore produce anche un file di metriche denominato `<Sample_ID>,roh_metrics.csv` che elenca il numero di ROH di grandi dimensioni e la percentuale di SNP nelle ROH di grandi dimensioni (> 3 MB).

## Metriche di stima della ploidia

Il Ploidy Estimator (Strumento di stima della ploidia) viene eseguito per impostazione predefinita. Il Ploidy Estimator (Strumento di stima della ploidia) utilizza le letture del mappatore/allineatore per calcolare la profondità di sequenziamento della copertura per ciascun autosoma e allosoma nel genoma umano. Il cariotipo sessuale del campione viene quindi stimato utilizzando i rapporti tra le coperture mediane dei cromosomi sessuali e la copertura mediana autosomica. XX o XY e CONCORDANT (CONFORME), DISCORDANT (NON CONFORME) o ND (Non determinato) rispetto ai dati dei campioni forniti sono riportati nel report consolidato. I risultati dettagliati, compresa ogni copertura mediana normalizzata per-contig, sono riportati nel file `<Sample_ID>.ploidy_estimation_metrics.csv`.

## File FASTQ

FASTQ (\*.fastq.gz, \*.fastq.ora) è un formato file di testo che contiene le identificazioni delle basi e i valori qualitativi per ogni lettura. Ogni file contiene le informazioni seguenti:

- Identificatore del campione
- La sequenza
- Un segno più (+)
- I punteggi qualitativi su scala Phred in un formato codificato ASCII + 33

Il software genera un file FASTQ per ogni campione, lettura e corsia. Ad esempio, per ogni campione in una corsa paired-end, il software genera due file FASTQ: uno per Read 1 (Lettura 1) e uno per Read 2 (Lettura 2). Oltre a questi file FASTQ dei campioni, il software genera due file FASTQ per corsia contenenti tutti i campioni sconosciuti. I file FASTQ per Index Read 1 (Lettura indice 1) e Index Read 2 (Lettura indice 2) non vengono generati perché la sequenza è inclusa nell'intestazione di ogni immissione FASTQ. Il formato del nome del file è costituito dai campi specificati nel foglio campioni e utilizza il formato di denominazione del file `<Sample_ID>_S#_L00#_R#_001.fastq.gz`

I file FASTQ vengono salvati nella directory `<Sample_ID>/Conversion`. Nella directory FASTQ della cartella di analisi è possibile trovare la directory Logs (Registri) con i registri di conversione da BCL a FASTQ e la directory Reports (Report) che contiene vari file di metriche di lettura e `SampleSheet.csv` utilizzato per la conversione FASTQ. I file FASTQ da letture indeterminate si trovano nella directory `Undetermined/Conversion` (Indeterminato/Conversione) della cartella Analysis (Analisi).

L'identificatore del campione è formattato nel seguente modo:

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber
```

Esempio:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
```

```
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
```

```
+
```

```
<>;##=><9=AAAAAAAAAAAA9#:<#<;<<<?????#=#
```

## File CRAM

I file della Mappa di allineamento orientata al riferimento compresso (CRAM) (\*.cram) sono memorizzati nella directory `<Sample_ID>/Analysis` e contengono intestazioni e record di allineamento relativi al file di riferimento genomico utilizzato durante l'allineamento. Il percorso del file di riferimento è indicato nel file `<Sample_ID>/Analysis/<Sample_ID>-replay.json`, come parametro `--ht-reference`, per impostazione predefinita impostato su `hg38.fa`.

I file CRAM contengono una sezione di intestazione e una sezione di allineamento:

- **Header** (Intestazione): contiene le informazioni sull'intero file, come il nome del campione, la lunghezza del campione e il metodo di allineamento. Gli allineamenti nella sezione allineamenti sono associati a informazioni specifiche nella sezione intestazione.
- **Alignments** (Allineamenti): contiene il nome della lettura, la sequenza della lettura, la qualità della lettura, le informazioni sull'allineamento e le tag personalizzate. Il nome della lettura include il cromosoma, la coordinata iniziale, la qualità dell'allineamento e la stringa del descrittore di corrispondenza.

La sezione degli allineamenti include le seguenti informazioni per ogni lettura o accoppiamento di letture:

- AS: Qualità dell'allineamento a estremità accoppiate.
- RG: Gruppo di lettura, che indica il numero di letture per un campione specifico.

- BC: Etichetta Barcode, che indica l'ID del campione demultiplexato associato alla lettura.
- SM: Qualità dell'allineamento a estremità singola.
- XC: Stringa del descrittore dell'accoppiamento.
- XN: Tag del nome dell'amplicone, che registra l'ID dell'amplicone associato con la lettura.

Per visualizzare i record di allineamento, `samtools` può essere utilizzato come `samtools view --reference <path_to_reference_folder>/hg38.fa <Sample_ID>.cram`.

Vengono generati anche un file indice e un file checksum.

# Assistenza tecnica

Per ricevere assistenza tecnica, contattare l'Assistenza tecnica Illumina.

**Sito web:** [www.illumina.com](http://www.illumina.com)

**E-mail:** [techsupport@illumina.com](mailto:techsupport@illumina.com)

**Schede dei dati di sicurezza (SDS):** sono disponibili sul sito web Illumina all'indirizzo [support.illumina.com/sds.html](http://support.illumina.com/sds.html).

**Documentazione sul prodotto:** disponibile per il download all'indirizzo [support.illumina.com](http://support.illumina.com).



Illumina, Inc.  
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  
The Netherlands

**Sponsor australiano**

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

PER USO DIAGNOSTICO IN VITRO.

© 2024 Illumina, Inc. Tutti i diritti riservati.

**illumina**<sup>®</sup>