

illumina®

# TruSight Whole Genome Analysis Application

Documentația produsului

PROPRIETATEA COMPANIEI ILLUMINA

Document nr. 200049931 v00

Aprilie 2024

A SE UTILIZA LA DIAGNOSTICAREA IN VITRO.

## Istoricul versiunilor

Document	Data	Descrierea modificării
Document nr. 200049931 v00	Aprilie 2024	Versiunea inițială.

Prezentul document și conținutul acestuia constituie proprietatea Illumina, Inc. și a afiliațiilor săi („Illumina”) și sunt destinate exclusiv pentru utilizarea contractuală de către client în legătură cu folosirea produsului sau produselor descrise în prezentul document și în niciun alt scop. Acest document și conținutul său nu trebuie utilizate sau distribuite pentru niciun alt scop și/sau nici comunicate, divulgate sau reproduse în orice alt mod și în orice formă fără consimțământul prealabil acordat în scris de Illumina. Illumina nu transmite, în temeiul brevetelor sale, al mărcilor sale comerciale, al drepturilor sale de autor sau în temeiul dreptului comun, nicio licență și nici drepturi similare ale oricăror terți prin acest document.

Instrucțiunile din acest document trebuie respectate în mod strict și explicit de către personalul calificat și corespunzător instruit pentru a asigura utilizarea corespunzătoare și în siguranță a produsului descris/produselor descrise în acest document. Înainte de utilizarea acestui produs/acestor produse, întreg conținutul acestui document trebuie citit și înțeles în întregime.

NERESPECTAREA OBLIGAȚIEI DE A CITI COMPLET ȘI DE A RESPECTA ÎN MOD EXPLICIT TOATE INSTRUCȚIUNILE CUPRINSE ÎN PREZENTUL DOCUMENT POATE DUCE LA DETERIORAREA PRODUSULUI SAU PRODUSELOR, LA VĂTĂMAREA PERSOANELOR, INCLUSIV A UTILIZATORILOR SAU A ALTOR PERSOANE ȘI LA DAUNE ALE ALTOR PROPRIETĂȚI ȘI VA ANULA ORICE GARANȚIE APLICABILĂ PRODUSULUI SAU PRODUSELOR.

ILLUMINA NU ÎȘI ASUMĂ NICIO RĂSPUNDERE CARE DECURGE DIN UTILIZAREA INADECVATĂ A PRODUSULUI SAU PRODUSELOR DESCRISE ÎN PREZENTUL DOCUMENT (INCLUSIV A COMPONENTELOR SAU SOFTWARE-ULUI ACESTORA).

© 2024 Illumina, Inc. Toate drepturile rezervate.

Toate mărcile comerciale sunt proprietatea Illumina, Inc. sau a proprietarilor lor respectivi. Pentru informații specifice privind mărcile comerciale, consultați [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).

# Cuprins

Istoricul versiunilor .....	ii
Prezentare generală .....	1
Noțiuni de bază .....	1
Setări .....	2
Rulare creare .....	4
Rulare revizie .....	5
Retrimiteri analiză în coadă .....	6
Vizualizare rulare și rezultate .....	7
Rezumat fișier de ieșire .....	8
Informații raport QC .....	8
Informații despre fișierul de definiții al variantei .....	12
Fișiere FASTQ .....	20
Fișiere CRAM .....	21
<b>Asistență tehnică .....</b>	<b>23</b>

# Prezentare generală

TruSight Whole Genome Analysis Application este utilizată pentru a planifica secvențierea executărilor pentru TruSight Whole Genome și a iniția automat analiza după finalizarea executării. Analiza include demultiplexarea, generarea FASTQ, maparea citirii, alinierea la genomul de referință uman GrCh38/hg38 cu grafic activat și definirea variantei utilizând Illumina DRAGEN Server pentru NovaSeq 6000Dx.

În diferite etape ale fluxului de lucru pentru analiză, aplicația efectuează controlul calității (QC) în conformitate cu parametrii de secvențiere definiți, FASTQ și biblioteca de probe și generează rapoarte cu rezultatele. Pentru probele care parcurg toți pașii CC, aplicația generează fișiere de ieșire de suport pentru utilizare în aplicațiile liniei germinale din aval.

TruSight Whole Genome Analysis Application execută definatorii variantei DRAGEN, inclusiv definatorul de variante mici, definatorul de numere de copii (CNV) și detectarea repetată a extinderii cu ExpansionHunter.

Aplicația efectuează, de asemenea, adnotarea unui nivel de încredere scăzut, intermediar sau ridicat pentru variantele mici și include această adnotare în fișierul de ieșire.

## Noțiuni de bază

Asigurați-vă că TruSight Whole Genome Analysis Application este instalat pe NovaSeq 6000Dx Instrument care va fi utilizat pentru secvențiere ca parte a TruSight Whole Genome. Aplicațiile instalate pot fi găsite pe ecranul Aplicații pe NovaSeq 6000Dx Instrument sau în Illumina Run Manager utilizând un browser pe un computer din rețea. Pentru asistență privind programarea instalării, contactați reprezentantul local Illumina pe teren.

### Cerințe privind depozitarea plasmei

Consultați Documentația produsului NovaSeq 6000Dx (document nr. 200010105) și Documentația produsului DRAGEN Server for NovaSeq 6000Dx (document nr. 200014171) pentru mai multe informații despre ieșirea și stocarea datelor.

Datele de ieșire TruSight Whole Genome Analysis Application sunt transmise într-un folder de rulare și într-un folder de analiză din spațiul de stocare extern. Cerințele minime de stocare pot fi approximate de la dimensiunea ieșirii datelor în fiecare folder pentru o singură executare de secvențiere prezentată mai jos.

Configurație	Rulare folder (GB)	Folderul de analiză (GB)
Celulă de flux S2 (6 probe)	~430	350
Celulă de flux S4 (16 probe)	~1110	890

## Ora aproximativă a analizei

Analiza începe automat după finalizarea unei rulări secvențiale și are loc secvențial pe probele din cadrul unei rulări. Fișierele de ieșire a datelor vor fi disponibile în spațiul de stocare extern după finalizarea analizei pentru toate probele dintr-o rulare, iar transferul copiei către spațiul de stocare extern este finalizat. Când se începe o executare de secvențiere atât pe partea A, cât și pe partea B în același timp, secvențierea se va efectua concomitent. Analiza acestor rulări de secvențiere va fi efectuată secvențial de către TruSight Whole Genome Analysis Application după finalizarea secvențierii. Rularea care finalizează secvențierea și transferul mai întâi va fi analizată mai întâi. A doua executare de secvențiere va fi transferată și pusă în așteptare pentru analiză după finalizarea primei analize. Consultați [Vizualizare rulare și rezultate la pagina 7](#) pentru modul de determinare a stării rulărilor active sau eșuate.

Timpul aproximativ până când rezultatele analizei sunt disponibile după finalizarea secvențierii este afișat mai jos pentru situația în care partea A și partea B sunt încărcate simultan cu aceeași configurație.

Configurație	Rulare analiză 1 (ore)	Rulare analiză 2 (ore)
Celulă de flux S2 (6 probe)	~12	~24
Celulă de flux S4 (16 probe)	~24	~48

## Setări

Selectați TruSight Whole Genome Analysis Application pe ecranul Application (Aplicații) pentru a vizualiza configurația curentă și pentru a modifica permisiunile.

### Configurație

Ecranul Configuration (Configurație) afișează următoarele setări ale aplicației:

- **Nume aplicație**
- **Versiune aplicație**
- **DRAGEN Versiune**
- **Versiune RTA**
- **Data versiunii**
- **Organizație**
- **Identificator dispozitiv**
- **Identificator producție**
- **Seturi de pregătire a bibliotecii** — afișează setul de pregătire a bibliotecii. Această setare nu poate fi modificată.
- **Seturi adaptoare index** — Afișează seturile de adaptoare index disponibile pentru utilizare.

- **Citiri indexuri**
- **Tip citire**
- **Lungimi index**
- **Lungimi de citire** — Lungimile de citire sunt setate în mod implicit atunci când este selectat setul de indexuri. Această setare nu poate fi modificată.

## Permisiuni

Administratorul desemnat are acces la Permisiuni și poate utiliza casetele de selectare de pe ecranul Permisiuni pentru a gestiona accesul utilizatorilor pentru TruSight Whole Genome Analysis Application.

Pentru mai multe informații privind permisiunile și gestionarea utilizatorilor, consultați secțiunea Configurare sistem din Documentația produsului NovaSeq 6000Dx (document nr. 200010105).

# Rulare creare

Creați rulări noi în modul IVD fie pe instrument, fie accesând Illumina Run Manager (IRM), folosind un browser de pe un computer din rețea. Pentru a accesa instrumentul de la distanță, utilizați informațiile despre adresă și contul de utilizator furnizate de reprezentantul dvs. Illumina. Consultați Documentația produsului NovaSeq 6000Dx (document nr. 200010105) pentru mai multe informații.

Crearea rulării este metoda recomandată pentru planificarea rulării. Nu se recomandă importarea fișei de probe. Leșirea fișierelor fișă probe din folderele de rulare și analiză nu este adecvată pentru import în timpul planificării rulării.

## Creare rulări

1. Din ecranul Runs (Rulări), selectați **Create Run** (Creare rulare).
2. Selectați TruSight Whole Genome Analysis Application, apoi selectați **Next** (Înainte).
3. Pe ecranul Run Settings (Setări rulare), introduceți un nume de rulare. Numele rulării identifică rularea din secvențiere prin analiză.
4. [Opțional] Introduceți o descriere de rulare care să ajute la identificarea ulterioară a rulării. Kitul de pregătire a bibliotecii este setat în mod implicit ca TruSight Whole Genome și nu poate fi modificat.
5. Selectați setul de indexuri TruSight Whole Genome dorit din meniul derulant **Index Adapter Kit** (Set adaptor pentru indexuri).  
Lungimea de citire va fi setată în mod implicit și nu poate fi modificată. (Citirea 1 și 2 utilizează 151 de cicluri; indicii 1 și 2 utilizează 10 cicluri).
6. Introduceți un ID de eprubetă din bibliotecă (format recomandat ca DX1234567-LIB ), apoi selectați **Next** (Înainte).  
Dacă nu este specificat niciun ID de eprubetă din bibliotecă în acest pas, rularea planificată va trebui selectată înainte de încărcarea consumabilelor de secvențiere. Dacă ID-ul incorect al eprubetei din bibliotecă este introdus în acest pas, rularea planificată trebuie corectată înainte de încărcarea consumabilelor. Consultați [Rulare revizie la pagina 5](#) pentru protocolul de corectare a rulării când sunteți gata să încărcați consumabile.
7. În ecranele Sample Data (Date probă) și Sample Settings (Setări probă), vor fi introduse informații despre probă. Datele probei pot fi introduse manual sau importând un fișier de date ale probei. ID-ul probei trebuie să fie unic pentru fiecare probă și poate conține numai caractere alfanumerice, caractere de subliniere și cratime. Nu includeți spații. Poziția godeului se referă la godeul în formatul de la A01 la H04 al plăcii index. Informațiile despre secvența indicelui vor fi populate automat atunci când este introdusă poziția godeului pe placa indice. Sexul trebuie introdus ca Masculin, Feminin sau Necunoscut. ID-ul plăcii de bibliotecă și ID-ul godeului din bibliotecă (de ex., formatul A01) sunt câmpuri obligatorii.



- Pentru a introduce manual datele probei, adăugați rânduri (până la un total de 6 pentru celula de flux S2 sau 16 pentru celula S4) și introduceți informațiile necesare în câmpurile ID probă și Poziție godeu. Informațiile pot fi copiate și lipite din Excel. Selectați **Next** (Înainte). Pe ecranul Sample Settings (Setări probă), introduceți ID placă bibliotecă, ID godeu bibliotecă și Sex. Selectați **Next** (Înainte).
  - Pentru a importa un fișier de date pentru probe, selectați **Import Samples** (Importare probe) și încărcați fișierul de date pentru probe. Informațiile vor fi populate automat în rânduri. Un șablon (\*.csv) este disponibil pentru descărcare pe acest ecran. Selectați **Next** (Înainte). În ecranul Sample Settings (Setări probă), informațiile vor fi populate în rânduri automat din fișierul cu date ale probei importat. Selectați **Next** (Înainte).
8. În ecranul Analysis Settings (Setări analiză), introduceți numele lotului înregistrat în timpul planificării lotului și executării.
  9. [Opțional] Selectați tipul celulei de flux, S2 sau S4.
  10. Confirmați sau debifați caseta de selectare pentru a Generare FASTQ-uri comprimate ORA, apoi selectați **Next** (Înainte).

**NOTĂ** TruSight Whole Genome Analysis Application generează FASTQ-uri comprimate ORA în mod implicit. Modificarea acestei setări va crește dimensiunea ieșirii finale de date.

11. Pe ecranul Run Review (Rulare revizuire), revizuiți informațiile introduse. Dacă nu sunt necesare modificări, selectați **Save** (Salvare). Dacă sunt necesare modificări, selectați **Back** (Înapoi) după cum este necesar pentru a reveni la ecranul corespunzător.



## ATENȚIE

TruSight Whole Genome a fost validat pentru 6 probe când se utilizează celula de flux NovaSeq 6000Dx S2 și pentru 16 probe când se utilizează celula de flux S4 NovaSeq 6000Dx. Asigurați-vă că introduceți numărul corect de probe pentru configurația selectată a celulei de flux.

## Rulare revizie

Dacă sunt necesare modificări după crearea rulării și înainte de încărcarea consumabilelor pentru secvențiere, revizuiți rulările în modul IVD fie pe instrument, fie accesând Illumina Run Manager (IRM) utilizând un browser de pe un computer din rețea.

1. Selectați **Runs** (Rulări).
2. Selectați numele rulării din fila Rulări planificate.
3. Selectați **Edit** (Editare).
4. Actualizați informațiile despre rulare sau probă după cum este necesar. De exemplu, introduceți sau corectați ID-ul eprubetei din bibliotecă pentru a se potrivi cu cel utilizat la finalizarea fluxului de lucru.

5. Selectați **Next** (Înainte) până la Run Review (Rulare revizuire).
6. Selectați **Save** (Salvare).
7. Selectați **Exit** (Ieșire).

Reveniți la Secvențiere în modul IVD pentru a repeta încărcarea consumabilelor. Rularea ar trebui să fie acum evidențiată automat.

Dacă actualizați ID-ul eprubetei din bibliotecă în timpul încărcării consumabilelor, reveniți la Run Selection (Selectare rulare) în software-ul de control și selectați **Refresh** (Reîmprospătare) pentru coloana asociată, A sau B. Rularea trebuie să fie acum evidențiată automat. Dacă nu, selectați **Back** (Înapoi) pentru a repeta Load consumables (Încărcare consumabile).

## Retrimiterere analiză în coadă

Consultați secțiunea Depanare din Prospectul TruSight Whole Genome (document nr. 200050132) pentru a determina ce tip de analiză a recuperării este cel mai adecvat.

### Retrimitererea analizei în coadă fără modificări

1. Selectați numele rulării complete pentru a vedea detaliile rulării.
2. Selectați **Requeue Analysis** (Retrimiterere analiză în coadă).
3. Selectați **Requeue Analysis with no changes** (Retrimitererea analizei în coadă fără modificări).
4. Furnizați detalii în câmpul Reanalysis Reason (Motiv reanaliză).
5. Selectați **Requeue Analysis** (Retrimiterere analiză în coadă).
6. Ieșiți din pagină și navigați la pagina Rulări active pentru a confirma că retrimitererea analizei în coadă este în curs.

### Retrimitererea analizei în coadă cu modificări

1. Selectați numele rulării complete pentru a vedea detaliile rulării.
2. Selectați **Requeue Analysis** (Retrimiterere analiză în coadă).
3. Selectați **Edit run settings** (Editarea setărilor rulării) și **Requeue Analysis** (Retrimitererea analizei în coadă).
4. Furnizați detalii în câmpul Reanalysis Reason (Motiv reanaliză).
5. Selectați **Requeue Analysis** (Retrimiterere analiză în coadă).
6. Confirmați sau actualizați Run Settings (Setări rulare), apoi selectați **Next** (Înainte).
7. Corectați informațiile probei după cum este necesar actualizând manual câmpurile sau selectați **Download Template** (Descărcare șablon) pentru a crea un fișier `sampledata.csv` cu informațiile curente. Corectați informațiile și ștergeți rândurile existente din fila Date probă înainte de a utiliza Import Samples (Importare probe) pentru a popula datele corectate ale probei.

8. Revizuiți informațiile din ecranul Run Review (Revizuire rulare) și selectați **Save** (Salvare) pentru a începe analiza.
9. Selectați **Exit** (Ieșire) și navigați la pagina Rulări active pentru a confirma că retrimiterrea analizei în coadă este în curs.

Folderul original de date despre rulare trebuie să fie prezent în locația de stocare externă specificată în Detalii executare pentru ca reanaliza să fie finalizată cu succes. Dacă reanaliza eșuează, asigurați-vă că rulara nu a fost mutată sau ștearsă.

## Vizualizare rulare și rezultate

1. Din ecranul principal Illumina Run Manager în modul IVD, selectați **Runs** (Rulări).
2. Din fila Completed Runs (Rulări finalizate), selectați numele Rulării.  
Această filă va afișa și rulările care s-au finalizate din cauza eșecului secvențierii, transferului de date sau analizei. Active runs (Rulări active) și starea acestora sunt afișate în fila Active Runs (Rulări active). Consultați Documentația produsului NovaSeq 6000Dx (document nr. 200010105) pentru mai multe informații.
3. Selectați numele Run (Rulare) în fila Completed Runs (Rulări finalizate) pentru a vizualiza Run Details (Detalii rulare) și Results for the path to the Analysis Output Folder (Rezultate pentru calea către Folderul de ieșire analiză).  
Pentru rulări eșuate, revizuiți Starea pentru fiecare pas și apoi consultați secțiunea Depanare din Prospectul TruSight Whole Genome (document nr. 200050132).
4. Navigați la folderul de analiză de pe unitatea locală și deschideți Raportul consolidat pentru a revizui rezultatul SUCCES/EȘEC pentru fiecare pas al CC după cum urmează:
  - Pentru CC al rulării secvențierii, consultați Rezumatul rezultatelor CC al secvențierii
  - Pentru controlul calității FASTQ pentru fiecare probă din rulare, consultați Rezumatul rezultatelor controlului calității FASTQ
  - Pentru controlul calității CC pentru fiecare probă din rulare, consultați Rezumatul rezultatelor controlului calității bibliotecii de probeDacă se observă un rezultat EȘEC, notați pasul CC și consultați secțiunea Depanare din Prospectul TruSight Whole Genome (document nr. 200050132).

## Rezumat fișier de ieșire

TruSight Whole Genome Analysis Application salvează următoarele fișiere de ieșire principale. Consultați secțiunile cu informații despre fișiere de mai jos pentru localizarea fișierelor de ieșire principale.

Rulările și probele care nu îndeplinesc criteriile de validitate nu produc fișiere CRAM, ROH sau \*genome.vcf.

Fișier de ieșire	Descriere
Raport consolidat (*.csv)	Conține indicatori de calitate utilizați pentru validitatea rulării (inclusiv randamentul total și Q30), indicatori de validitate a probelor (inclusiv randamentul FASTQ), indicatori de CC al bibliotecii și indicatori doar pentru informații (FIO) pentru toate probele din rulare.
Raport probă (*.csv)	Conține rezultate ale secvențierii CC, FASTQ și CC al bibliotecii de probe. Raportul conține, de asemenea, concordanța de ploidie și parametri FIO pentru proba individuală, precum și executarea de secvențiere asociată.
Variantă mică și VCF mSNV (*.adnotat.hard-filtrat-gvcf.gz)	Conține informații despre definițiile variantelor pentru variantele mici (SNV, indels) și SNV mitocondriale.
CNV VCF (*.cnv.vcf.gz)	Conține informații despre definițiile variantelor pentru câștigurile și pierderile numărului de copii.
Repetă VCF (*.repeats.vcf.gz)	Conține informații de definire variate despre extinderile STR și SMN1.
BAZĂ ROH (*.roh.bed)	Conține informații pentru regiunile cu homozigotate.
FASTQ (*.fastq.gz sau *.fastq.ora)	Fișiere intermediare care conțin definițiile bazelor evaluate din punct de vedere al calității. Fișierele FASTQ constituie principala introducere de date pentru etapa de aliniere. Dacă este selectată compresia ORA, numele fișierului reflectă acest lucru.
CRAM aliniere (*.cram)	Conține citiri aliniat pentru o probă dată.

## Informații raport QC

Raportul consolidat <<RunID>>\_Consolidated\_Report.csv se află în directorul TruSightWholeGenomeAnalysis\_x.x.x\_run-complete și conține informații despre parametrii de

calitate utilizați pentru a trece sau a eșua probele în diferite etape de analiză. Rapoartele probelor individuale <<Sample\_ID>>\_Sample\_Report.csv pot fi găsite în directorul <Sample\_ID> din directorul TruSightWholeGenomeAnalysis\_x.x.x\_run-complete.

Antetele raportului includ următoarele informații despre rulare: versiunea aplicației, numele lotului, ID-ul fondului de eprubete al bibliotecii, numele executării secvențierii, ID-ul executării secvențierii și tipul celulei de flux. Tabelele următoare descriu informațiile incluse în Raportul consolidat. Raportul probei individuale include aceleași informații, cu excepția măsurătorilor demultiplex.

Tabelul 1 Metricile CC pentru secvențiere

Valori	Specificații	Descriere
Randament total neindexat (GB)	Nu se aplică	Nicio specificație, deoarece volumele de producție mai mici pot duce la trecerea bibliotecilor de probe. Așteptați-vă la $\geq 3000$ Gbp pentru S4 și $\geq 1000$ GB pentru celula de flux S2.
Total % $\geq$ Q30	$\geq 85$	Măsurarea calității bazei la nivelul rulării. Specificația minimă este setată întrucât rulările %Q30 prea scăzute nu vor trece de bazele Q30 din Controlul calității bibliotecii de probe.
Rezumat rezultat CC secvențiere	PASS (Succes) sau FAIL (Eșec)	Pentru eroare CC de secvențiere, consultați secțiunea Depanare din Prospectul TruSight Whole Genome (document nr. 200050132).

Tabelul 2 Măsurători de demultiplexare

Valori	Specificații	Descriere
Procentul citirilor identificate	Nu se aplică	Fracția totală a citirilor filtrului de trecere din rulare care au fost alocate probelor în timpul demultiplexării.
Procent CV	Nu se aplică	Oferă o măsurare a uniformității citirilor demultiplexate pentru fiecare pereche de indici din timpul rulării. Așteptați-vă la $< 25\%$ pentru rulările fără erori ale rezultatelor CC FASTQ.

Tabelul 3 Indicatori de CC FASTQ

Valori	Specificații	Descriere
Randament per probă (bps)	$\geq 90.000.000.000$	Valoarea minimă este setată să fie echivalentă cu ~26 de ori acoperirea autozomală medie cu cea a probelor pentru trierea bibliotecilor de probe care nu trec de controlul calității bibliotecii pentru a reduce timpul de analiză.

Valori	Specificații	Descriere
Rezumatul rezultatelor controlului calității FASTQ	PASS (Succes) sau FAIL (Eșec)	Pentru eroare CC FASTQ, consultați secțiunea Depanare din Prospectul TruSight Whole Genome (document nr. 200050132).

Tabelul 4 Sample Library QC Metrics (Valori CC bibliotecă de probe)

Valori	Specificații	Descriere
Acoperire autozomală medie	$\geq 35$	Acoperire medie printre autozomi. Specificațiile minime sunt stabilite pentru a asigura performanța analitică.
Procentajul autozomilor cu acoperire mai mare de 20X	$\geq 93,94$	Măsurarea uniformității acoperirii care detectează probleme nu este legată neapărat de eroarea sistematică GC. Specificațiile minime sunt stabilite pentru a asigura performanța analitică.
Acoperire normalizată la compartimente GC de la 60% la 79%	$0,82 \leq x \leq 1,13$	Măsura uniformității acoperirii care detectează eroarea sistematică GC, în special o pierdere a acoperirii în zonele genomului cu % GC mai mare și compoziție de bază mai mică % AT. Specificațiile minime și maxime sunt stabilite pentru a asigura performanța analitică.
Acoperire normalizată la compartimente GC de la 20 % la 39%	$0,97 \leq x \leq 1,06$	Măsura uniformității acoperirii care detectează eroarea sistematică GC, în special o pierdere a acoperirii în zonele genomului cu % GC mai mică și compoziție de bază mai mare % AT. Specificațiile minime și maxime sunt stabilite pentru a asigura performanța analitică.
Acoperire mitocondrială medie	$\geq 500$	Acoperirea cromozomului mitocondrial. Specificația minimă este setată pentru a asigura limita de detecție SNV mitocondrial.
Procentul bazelor Q30	$\geq 85$	Măsurarea calității bazei. Specificațiile minime sunt stabilite pentru a asigura performanța analitică.
Contaminare estimată a probei	$\leq 0,005$	Detectează citirile contaminante din alte probe. Specificația maximă este setată pentru a asigura limita de detecție SNV mitocondrială (tipul de variantă cu cea mai mare sensibilitate la contaminare).
Rezumat rezultat CC bibliotecă de probe	PASS (Succes) sau FAIL (Eșec)	Pentru eroare CC a bibliotecii de probe, consultați secțiunea Depanare din Prospectul TruSight Whole Genome (document nr. 200050132).

Tabelul 5 Măsurători CC pentru ploidie

Valori	Specificații	Descriere
A furnizat ploidie cromozomială sexuală	Nu se aplică	Sexul furnizat de operator în timpul creării rulării (Feminin, Masculin, Necunoscut).
Estimarea ploidiei	Nu se aplică	Ploidie sexuală estimată de DRAGEN.
Rezultat ploid rezumat	CONCORDANT, DISCORDANT sau ND	CONCORDANT indică acordul între ploidia sexuală furnizată și cea estimată. ND indică sexul furnizat ca Necunoscut sau estimare diferită de XX sau XY. Pentru rezultatele DISCORDANT, a fost introdus un sex greșit în timpul creării rulării sau este posibil să fi avut loc schimbarea probei. Consultați secțiunea Depanare din Prospectul TruSight Whole Genome (document nr. 200050132).

Tabelul 6 Valori doar în scop informativ

Valori	Descriere
Introduceți lungimea mediană	Ținta este de 450 bp, dar se așteaptă variații prin ciclul de secvențiere și operator. Este acceptabil un interval de aproximativ 360 până la 550 bp. Funcționarea consecventă în afara acestui interval poate duce la o incidență mai mare a defectării probei.
Procente de citiri cartografiate	Procentajul citirilor care s-au aliniat la genomul de referință. Poate fi scăzut ca răspuns la contaminarea cu ADNg non-uman, calitatea slabă a probei sau lungimea prea mică a insertului, ducând la probleme de cartografiere.
Procente de citiri cu alinieri suplimentare	Procentul de citiri cu mapare care se împarte în diferite locații din genomul de referință.
Procent citiri marcate în duplicat	Așteptați-vă la < 20%. Poate fi ridicată dacă randamentul de pregătire a bibliotecii este scăzut sau dacă se cumulează mai puțin decât volumul necesar sau ca răspuns la problemele legate de secvențiere.
Procent de baze decupate moi Citire 1	Util în diagnosticarea cauzei principale pentru eșecul acoperirii autozomale medii.

Valori	Descriere
Procent de baze decupate moi Citire 2	Util în diagnosticarea cauzei principale pentru eșecul acoperirii autozomale medii.
Procente de baze tăiate Citire 1	Util în diagnosticarea cauzei principale pentru eșecul acoperirii autozomale medii.
Procente de baze tăiate Citire 2	Util în diagnosticarea cauzei principale pentru eșecul acoperirii autozomale medii.

## Informații despre fișierul de definiții al variantei

### Fișiere VCF

Fișierele în formatul de definiție a variantelor (\*.vcf) conțin informații despre variantele situate în poziții specifice într-un genom de referință și pot fi găsite în directorul <Sample\_ID>/Analysis.

Antetul fișierelor VCF include versiunea de format a fișierului VCF și versiunea definatorului de variante și enumeră adnotările folosite în restul fișierului. Ultimul rând din antet conține anteturile de coloană pentru liniile de date. Fiecare dintre liniile de date ale fișierului VCF conține informații despre o singură poziție de referință.

Toate fișierele VCF conțin un antet cu descrieri ale coloanelor de ieșire și ale datelor de definiție a variantelor în coloanele etichetate ca CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO, FORMAT, SAMPLE. Definițiile valorilor coloanei pot varia între definatorii variantelor.

### Variantă mică și mSNV VCF

Rezultatul este salvat în fișierul <Sample\_ID>.adnotated.hard-filtered.gvcf.gz din directorul <Sample\_ID>/Analysis.

Un fișier VCF genomic (gVCF) conține informații despre variantele și pozițiile determinate a fi homozigote la genomul de referință. Pentru regiunile homozigote, fișierul gVCF include statistici care indică cât de bine acceptă citirile absența variantelor sau a alelelor alternative. Fișierul gVCF include o alelă artificială <NON\_REF>. Citirile care nu acceptă referința sau orice variante sunt alocate alelei <NON\_REF>. DRAGEN utilizează aceste citiri pentru a determina dacă poziția poate fi denumită ca referință homozigotă, spre deosebire de a rămâne nedefinită. Punctajul rezultat reprezintă nivelul de încredere crescut într-o definiție de referință homozigotă. În modul linie germinală, scorul este FORMAT/GQ.



DRAGEN asigură filtrarea variantei post-VCF pe baza adnotărilor prezente în înregistrările VCF. Filtrarea dură a variantei este descrisă mai jos. Totuși, datorită naturii algoritmilor DRAGEN, care încorporează ipoteza erorilor corelate din interiorul miezului definatorului variantei, conducta are capacitate îmbunătățite în distingerea variantelor reale de zgomot și, prin urmare, dependența de filtrarea post-VCF este redusă substanțial.

TruSight Whole Genome Analysis Application oferă adnotarea scorului de încredere și a nivelului de încredere pentru variantele mici care pot fi utilizate pentru a îmbunătăți și mai mult performanța. Adnotarea nivelului de încredere nu este un filtru de calitate și, prin urmare, nu este reflectată direct în starea calității definirilor variantelor. Prin urmare, este posibil să vedeți definirile variantelor care sunt totuși adnotate ca având încredere scăzută.

Tabelul 7 Anteturile fișierelor VCF

Antet	Descriere
CHROM	Cromozomul genomului de referință. Cromozomii apar în aceeași ordine ca în fișierul de referință FASTA.
POS	Poziția de bază unică a variantei din cromozomul de referință. Pentru variante mononucleotidice (SNV), această poziție este baza de referință cu varianta. Pentru indeli, această poziție este baza de referință imediat anterioară variantei.
ID	Întotdeauna .
REF	Genotipul de referință. De exemplu, ștergerea unui singur T este reprezentată ca TT de referință și T alternativ. O variantă de nucleotidă unică de la A la T este reprezentată ca A de referință și T alternativ.
ALT	Alelele care diferă față de citirea de referință. De exemplu, o inserție a unui singur T este reprezentată ca A de referință și AT alternativ. O variantă de nucleotidă unică de la A la T este reprezentată ca A de referință și T alternativ.
QUAL	Un scor de calitate pe scara Phred atribuit de definatorul de variante. Scorurile mai mari indică o încredere mai mare în variantă și o probabilitate mai mică de erori. Pentru un scor de calitate Q, probabilitatea estimată a unei erori este de 10 <sup>-Q</sup> (Q/10). De exemplu, setul de definiri Q30 are o rată de erori de 0,1 %. Multe definoare de variante atribuie scoruri de calitate pe baza modelelor lor statistice, care sunt mari în raport cu rata de erori observată.

Tabelul 8 Adnotările fișierelor VCF

Antet	Descriere
FILTER (FILTRU)	<ul style="list-style-type: none"> <li>• <b>PASS</b> (Succes): Toate filtrele au avut succes.</li> <li>• <b>DRAGENSnpHardQUAL</b>—Setați dacă este adevărat:QUAL &lt; 10,41</li> <li>• <b>DRAGENIndelHardQUAL</b>—Setați dacă este adevărat:QUAL &lt; 7,83</li> <li>• <b>LowDepth</b>—Setați dacă este adevărat:DP &lt;= 1</li> <li>• <b>LowGQ</b>—Setați dacă este adevărat:GQ = 0</li> <li>• <b>PloidyConflict</b>—Definirea genotipului de la definatorul de variante nu este în concordanță cu ploidia cromozomilor</li> <li>• <b>base_quality</b>—Locație eliminată prin filtrare deoarece calitatea mediană a bazei pentru citirile ALT în acest locus nu atinge pragul</li> <li>• <b>filtered_reads</b>—Locație eliminată prin filtrare deoarece o fracțiune prea mare de citiri a fost eliminată prin filtrare</li> <li>• <b>fragment_length</b>—Locație eliminată prin filtrare pentru că diferența absolută dintre lungimea mediană a fragmentului citirilor ALT și lungimea mediană a fragmentului de citiri REF în acest locus depășește pragul</li> <li>• <b>low_af</b>—Frecvența alelelor nu îndeplinește pragul</li> <li>• <b>low_depth</b>—Locație eliminată prin filtrare deoarece adâncimea de citire este prea scăzută</li> <li>• <b>low_frac_info_reads</b>—Locație eliminată prin filtrare deoarece fracțiunea de citiri informative este sub prag</li> <li>• <b>low_normal_depth</b>—Locație eliminată prin filtrare deoarece adâncimea de citire a probei normale este prea scăzută</li> <li>• <b>long_indel</b>—Locație eliminată prin filtrare deoarece lungimea indelului este prea mare</li> <li>• <b>mapping_quality</b>—Locație eliminată prin filtrare deoarece calitatea mediană a mapării pentru citirile ALT în acest locus nu atinge pragul</li> <li>• <b>multialelic</b>—Locație eliminată prin filtrare deoarece mai mult de două alele ALT trec de LOD tumorală</li> <li>• <b>non_homref_normal</b>—Locație eliminată prin filtrare deoarece genotipul probei normale nu este o referință homozigotă</li> <li>• <b>no_reliable_supporting_read</b>—Locație eliminată prin filtrare deoarece nu există o citire somatică de susținere fiabilă</li> <li>• <b>panel_of_normals</b>—Văzut în cel puțin o probă din panoul de vcf cu probe normale</li> <li>• <b>read_position</b>—Locație eliminată prin filtrare deoarece mediana distanțelor dintre începutul/sfârșitul citirii și acest locus este sub prag</li> <li>• <b>RMxNRepeatRegion</b>—Locație eliminată prin filtrare deoarece întreaga sau o parte a alelei variantei este o repetare a referinței</li> <li>• <b>strand_artifact</b>—Locație eliminată prin filtrare din cauza unui decalaj sever al catenelor</li> </ul>

Antet	Descriere
FILTER (FILTRU)	<ul style="list-style-type: none"> <li>• <b>str_contraction</b>—Locație eliminată prin filtrare din cauza erorii PCR suspectate, unde alela ALT este cu o unitate de repetare mai mică decât referința</li> <li>• <b>too_few_supporting_reads</b>—Locație eliminată prin filtrare deoarece există prea puține citiri de susținere în proba tumorală</li> <li>• <b>weak_evidence</b>—Scorul variantei somatice nu atinge pragul</li> </ul>
INFO (INFORMAȚII)	<ul style="list-style-type: none"> <li>• <b>DB</b>—Calitatea de membru dbSNP.</li> <li>• <b>FS</b>—Valoare p pe scara Phred, utilizând testul exact al lui Fisher pentru a detecta decalajul catenelor.</li> <li>• <b>QD</b>—Încrederea/calitatea variantelor în funcție de profunzime.</li> <li>• <b>R2_5P_bias</b>—Scor bazat pe decalajul dintre corespondenți și distanța de la capătul 5 al catenei.</li> <li>• <b>SOR</b>—Raportul probabilităților simetrice a tabelului de contingență 2x2 pentru detectarea decalajului dintre catene.</li> <li>• <b>DP</b>—Profunzimea aproximativă a citirilor (informative și neinformative); este posibil ca unele citiri să fi fost eliminate prin filtrare în baza mapq etc.</li> <li>• <b>END</b>—Poziție de final a intervalului.</li> <li>• <b>FractionInformativeReads</b>—Frația de citiri informative din totalul citirilor.</li> <li>• <b>MQ</b>—Calitatea mapării RMS.</li> <li>• <b>MQRankSum</b>—Scorul Z la testul sumei rangurilor Wilcoxon pentru calitățile de mapare Alt vs. Ref.</li> <li>• <b>ReadPosRankSum</b>—Scorul Z la testul sumei rangurilor Wilcoxon pentru eroarea sistematică de poziție Alt vs. Ref.</li> <li>• <b>SOMATIC</b>—Cel puțin o variantă în această poziție este somatică.</li> <li>• <b>ILENS</b>—Lungimi indel pentru fiecare variantă ALT.</li> <li>• <b>SCORE</b>—Scor de încredere pentru fiecare tip de variantă prezentă la centru ca (tip de variantă):(scor de încredere).</li> <li>• <b>TIER</b>—Nivel de încredere pentru fiecare tip de variantă prezentă la centru ca (tip de variantă):(nivel de încredere).</li> </ul>

Antet	Descriere
FORMAT	<p>Coloana FORMAT enumeră câmpuri separate de două puncte, de exemplu GT:GQ.</p> <ul style="list-style-type: none"> <li>• <b>AD</b>—Profunzimi alelice (numărând doar citirile informative din totalul citirilor) pentru alelele de REF și ALT în ordinea enumerată.</li> <li>• <b>AF</b>—Frații de alele pentru alelele ALT în ordinea enumerată.</li> <li>• <b>DP</b>—Profunzimea aproximativă a citirilor (sunt filtrate citirile cu MQ=255 sau cu perechi inadecvate).</li> <li>• <b>F1R2</b>—Numărul de citiri în orientarea perechii F1R2 care susține fiecare alelă.</li> <li>• <b>F2R1</b>—Numărul de citiri în orientarea perechii F2R1 care susține fiecare alelă.</li> <li>• <b>GP</b>—Probabilități posterioare pe scara Phred pentru genotipuri, astfel cum este definit în specificația VCF.</li> <li>• <b>GQ</b>—Calitatea genotipului.</li> <li>• <b>GT</b>—Genotip.</li> <li>• <b>ICNT</b>—Număr de citiri informative INDEL pe baza modelului de încredere de referință.</li> <li>• <b>MB</b>—Statisticile componentelor pentru fiecare probă pentru a detecta decalajul dintre corespondenți.</li> <li>• <b>MIN_DP</b>—DP minim observat în blocul GVCF.</li> <li>• <b>PL</b>—Probabilitățile normalizate pe scara Phred pentru genotipuri, astfel cum este definit în specificația VCF.</li> <li>• <b>PRI</b>—Probabilități posterioare pe scara Phred pentru genotipuri.</li> <li>• <b>PS</b>—Informații privind ID-ul de etapizare fizică, unde fiecare ID unic dintr-o probă dată (dar nu la nivelul tuturor probelor) conectează înregistrările dintr-un grup de etapizare.</li> <li>• <b>SB</b>—Statisticile componentelor pentru fiecare probă care cuprind testul exact al lui Fisher pentru a detecta decalajul catenelor.</li> <li>• <b>SPL</b>—Probabilități normalizate, pe scala Phred pentru SNP pe baza modelului de încredere de referință.</li> <li>• <b>SQ</b>—Calitatea somatică.</li> </ul>
SAMPLE (PROBĂ)	Coloana Sample (Probă) prezintă valorile specificate în coloana FORMAT.

## Copiere număr varianta VCF

Etapa de numărare a țintelor este prima etapă de procesare pentru conducta CNV DRAGEN, producând `<Sample_ID>.target.counts.gz`, apoi se efectuează corecția de eroare sistematică GC, generând un fișier `*.target.counts.gz corectat.gz`. Etapa de normalizare produce fișierul `*.tn.tsv.gz`. Software-ul DRAGEN găzduiește multe fișiere intermediare. `*.seg.called.merged` este fișierul defenirii finale care conține evenimentele de amplificare și deleție.

În plus față de fișierul segment, DRAGEN emite definițiile în format VCF standard. Rezultatul este salvat în <Sample\_ID>.cnv.vcf.gz în directorul <Sample\_ID>/Analysis.

Definiții ale coloanelor specifice definatorului CNV:

Coloana POS este poziția de pornire a variantei. Conform specificației VCF, dacă oricare dintre alelele ALT este o alelă simbolică, cum ar fi <DEL>, atunci baza de acoperire este necesară, iar POS denotă coordonatele bazei care precedă polimorfismul. Toate coordonatele din VCF se bazează pe 1.

Coloana ID este utilizată pentru a reprezenta evenimentul. Câmpul ID codifică tipul de eveniment și coordonatele evenimentului.

Coloana REF conține un N pentru toate evenimentele CNV.

Coloana ALT specifică tipul de eveniment CNV. Deoarece VCF conține numai evenimente CNV, se utilizează numai intrarea DEL sau DUP.

Coloana QUAL conține un scor de calitate estimat pentru evenimentul CNV, care este utilizat în filtrarea dură.

Coloana FILTER (filtru) conține PASS (Succes) dacă evenimentul CNV trece de toate filtrele, în caz contrar coloana conține numele filtrului eșuat.

Coloana INFO conține informații care reprezintă evenimentul. Intrarea REFLEN indică durata evenimentului. Intrarea SVTYPE este întotdeauna CNV. Intrarea END (sfârșit) indică poziția finală a evenimentului.

Câmpurile FORMAT sunt descrise în antet.

- GT—Genotip
- SM — Raportul de copii liniare al mediei segmentului
- CN — Numărul estimat de copii
- BC — Numărul de compartimente din regiune
- PE — Numărul de citiri de capăt asociate necorespunzător la punctele de întrerupere la început și la sfârșit

## Repetă VCF

ExpansionHunter efectuează o realiniere pe bază de secvențe a citirilor care provin din interiorul și din jurul fiecărei repetiții țintă. ExpansionHunter genotipează apoi lungimea repetiției în fiecare alelă pe baza acestor alinieri ale graficului.

Mai multe informații și analize sunt disponibile în următoarele documente ExpansionHunter:

- Dolzhenko et al., *Detection of long repeat expansions from PCR-free whole-genome sequence data* 2017
- Dolzhenko et al., *ExpansionHunter: Un instrument bazat pe secvențe pentru a analiza variația în regiunile scurte cu repetare în tandem* 2019

Catalogul variantelor TruSight Whole Genome Analysis Application de STR conține specificații privind repetițiile care cauzează boli localizate în genele 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 și TBP.

Rezultatele genotipării repetate sunt transmise sub forma unui fișier VCF separat, care furnizează lungimea fiecărei alele la fiecare repetare definibilă precizată în fișierul catalog cu specificații repetate. Numele fișierului este <Sample\_ID>.repeats.gz și poate fi găsit în directorul <Sample\_ID>/Analysis.

Unele coloane sunt specifice pentru repetarea definatorului de extindere:

Tabelul 9 Câmpuri VCF de bază

Câmp	Descriere
CHROM	Identificator cromozom
POS	Poziția primei baze înainte de regiunea repetată în referință
ID	Întotdeauna .
REF	Baza de referință la poziția POS
ALT	Lista alelelor repetate în format <STRn>. N este numărul de unități repetate:
QUAL	Întotdeauna .
FILTER	Filtrul LowDepth este aplicat atunci când adâncimea totală a locusului este sub 10x sau numărul de citiri care se întind pe unul sau ambele capete de întrerupere este sub 5.

Tabelul 10 Câmpuri INFO suplimentare

Câmp	Descriere
END	Poziția ultimei baze a regiunii repetate în referință
REF	Număr copie de referință
REPID	ID-ul variantei din catalogul variantei
RL	Lungime de referință în bp
RU	Unități repetate în orientarea de referință
VARID	ID-ul variantei din catalogul variantei

Tabelul 11 Câmpuri GENOTIP (per probă)

Câmp	Descriere
AD	Adâncimi alelice pentru alelele ref și alelele alt în ordinea listată
ADFL	Număr de citiri de flancare în concordanță cu alela

Câmp	Descriere
ADIR	Număr de citiri repetate conforme cu alela
ADSP	Număr de citiri care se întind în conformitate cu alela
DST	Rezultatele (+ detectate, - nedetectate, ? nedeterminate) ale testului reprezentat de variantă
GT	Genotip
LC	Acoperirea locusului
REPCI	Interval de încredere pentru REPCN
REPCN	Numărul de unități repetate acoperite de alelă
RPL	Rație de probabilitate logaritmică pentru prezența alelei de referință
SO	Tipul de citiri care susțin alela. Valorile pot fi <i>SPANNING</i> (întindere), <i>FLANKING</i> (flancare) sau <i>INREPEAT</i> (în repetare). Aceste valori indică dacă citirile se întind, sunt flancate sau sunt conținute complet în repetare.

Fișierul <Sample\_ID>.repeats.vcf.gz include rezultatul SMN împreună cu orice repetiții țintite. Rezultatul SMN este reprezentat ca o singură definiție SNV la poziția de afectare a matisării în SMN1 (NM\_000344.3:c.840C/T) cu starea atrofiei musculare spinale (AMS) în următoarele câmpuri personalizate.

Tabelul 12 Rezultate SMA în fișierul de ieșire repeats.vcf

Câmp	Descriere
VARID	SMN marchează definiția SMN.
GT	Definiție genotip în această poziție utilizând un model de genotip normal (diploid).
DST	Definiție de stare AMS: + indică detectat - indică nedetectat ? indică nedeterminat
AD	Număr total de citiri care acceptă alela C și T.
RPL	Raportul de probabilitate Log10 între modelele neafectate și cele afectate. Scorurile pozitive indică o probabilitate mai mare a modelului neafectat.

## BAZA ROH

Regiunile de homozigotitate (ROH) sunt detectate ca parte a definiției variantei mici. Definiția detectează și transmite rulările de homozigotitate din definițiile întregului genom pe cromozomi umani autozomali. Cromozomii sexuali sunt ignorați, cu excepția cazului în care cariotipul sexual al probei este XX, așa cum este determinat de Estimatorul de Ploidie. Ieșirea ROH permite instrumentelor din aval să

selechteze și să anticipeze consangvinitatea dintre părinții subiectului proband de testare.

O regiune este definită ca definiții consecutive ale variantelor de pe cromozom, fără un spațiu mare între aceste variante. Cu alte cuvinte, regiunile sunt rupte de cromozom sau de lacune mari fără definiții SNV. Dimensiunea distanței este setată la 3 Mbases.

Definitorul ROH produce un fișier de rezultat ROH denumit `<Sample_ID>.roh.bed` în directorul `<Sample_ID>/Analysis`. Fiecare rând reprezintă o regiune de homozigozitate. Fișierul bed conține următoarele coloane:

```
Cromozom Start Sfârșit Scor Număr homozigot Număr heterozigot
```

Unde

- Scorul este o funcție a numărului de variante homozigote și heterozigote, unde fiecare variantă homozigotă crește scorul cu 0,025, iar fiecare variantă heterozigotă reduce scorul cu 0,975.
- Pozițiile de pornire și de terminare sunt un interval pe bază de 0, pe jumătate deschis.
- #Homozigot este numărul de variante homozigote din regiune.
- #Heterozigot este numărul de variante heterozigote din regiune.

De asemenea, definitorul produce un fișier de măsurare denumit `<Sample_ID>.roh_metrics.csv` care listează numărul de ROH mari și procentajul de SNP în ROH mare (> 3 MB).

## Parametri de estimare a ploidiei

Estimatorul de Ploidie rulează în mod implicit. Estimatorul de Ploidie utilizează citiri de la mapator/aliniator pentru a calcula adâncimea de secvențiere a acoperirii pentru fiecare autozom și alozom din genomul uman. Cariotipul sexual al probei este apoi estimat utilizând raporturile dintre acoperirile cromozomului sexual median și acoperirea autozomală mediană. XX sau XY și CONCORDANT, DISCORDANT sau ND (Nedeterminat) în comparație cu mostrele de date furnizate sunt raportate în raportul consolidat. Rezultatele detaliate, inclusiv fiecare acoperire medie normalizată per cont, sunt raportate în fișierul `<Sample_ID>.ploidy_estimation_metrics.csv`.

## Fișiere FASTQ

FASTQ (\*.fastq.gz, \*.fastq.ora) este un format de fișiere bazat pe text, care conține definițiile bazelor și valorile de calitate per citire. Fiecare fișier conține următoarele informații:

- Identificatorul probei
- Secvența
- Un semn plus (+)
- Scorurile de calitate Phred într-un format codificat ASCII + 33

Software-ul generează un fișier FASTQ pentru fiecare probă, citire și bandă. De exemplu, pentru fiecare probă dintr-o executare cu două capete, software-ul generează două fișiere FASTQ: unul pentru Citirea 1 și unul pentru Citirea 2. Pe lângă aceste fișiere FASTQ de probă, software-ul generează două fișiere FASTQ pe fiecare bandă care conține toate probele necunoscute. Fișierele FASTQ pentru Citire index 1



și Citire index 2 nu sunt generate deoarece secvența este inclusă în antetul fiecărei intrări FASTQ. Formatul numelui fișierului este construit din câmpurile specificate în fișa de probe și utilizează formatul de denumire tip `<Sample_ID>_S#_L00#_R#_001.fastq.gz`

Fișierele FASTQ sunt salvate în directorul `<Sample_ID>/Conversion`. În directorul FASTQ al directorului de analiză, se poate găsi directorul Journale cu jurnalele de conversie BCL-to-FASTQ și directorul Rapoarte, care conține diverse fișiere de valori citite și `SampleSheet.csv` utilizate pentru conversia FASTQ. Fișierele FASTQ din citiri nedeterminate se găsesc în directorul `Nedeterminat/Conversie` al folderului Analiză.

Identificatorul probei are următorul format:

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

Example:

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

```
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
```

```
+
```

```
<>;##=><9=AAAAAAAAAAA9#:<#<;<<<????#=#
```

## Fișiere CRAM

Fișierele comprimate de tip hartă de aliniere orientată spre referință sau CRAM (\*.cram) sunt stocate în directorul `<Sample_ID>/Analysis` și conțin antete și înregistrări de aliniere în raport cu fișierul de referință genomic utilizat în timpul alinierii. Călea către fișierul de referință este enumerată în fișierul `<Sample_ID>/Analysis/<Sample_ID>-replay.json`, ca parametru `--ht-reference`, setat implicit la `hg38.fa`.

Fișierele CRAM conțin o secțiune antet și o secțiune de alinieri:

- **Header** (Antet) – conține informații despre întregul fișier, precum denumirea probei, lungimea probei și metoda de aliniere. Alinierea din secțiunea de alinieri sunt asociate cu informații specifice din secțiunea antet.
- **Alignments** (Alinieri) – conține denumirea citirii, secvența citirii, calitatea citirii, informații despre alinieri și etichete personalizate. Denumirea citirii include cromozomul, coordonata de începere, calitatea alinierilor și șirul descriptor al potrivirilor.

Secțiunea de alinieri include următoarele informații pentru fiecare citire sau pereche de citiri:

- AS: calitatea alinierilor cu perechi de baze împerecheate.
- RG: grupul de citiri, care indică numărul de citiri pentru o anumită probă.
- BC: etichetă cod de bare, care indică ID-ul probei demultiplexate asociat cu citirea.
- SM: calitatea alinierilor cu pereche de baze unică.
- XC: șirul descriptor al potrivirilor.
- XN: Etichetă denumire amplicon, care înregistrează ID-ul ampliconului asociat cu citirea.

Pentru a vizualiza înregistrările alinierii, pot fi utilizate `samtools` ca `samtools view --reference <path_to_reference_folder>/hg38.fa <Sample_ID>.cram`.

De asemenea, se generează un fișier index și un fișier checksum (sumă de control).

## Asistență tehnică

Pentru asistență tehnică, contactați departamentul Asistență tehnică al Illumina.

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

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

**Fișe cu date de securitate (SDS)** – disponibile pe site-ul web Illumina la adresa [support.illumina.com/sds.html](http://support.illumina.com/sds.html).

**Documentația produselor** – disponibilă pentru descărcare de pe [support.illumina.com](http://support.illumina.com).



Illumina, Inc.  
5200 Illumina Way  
San Diego, California 92122 S.U.A.  
+1.800.809.ILMN (4566)  
+1.858.202.4566 (în afara Americii de Nord)  
techsupport@illumina.com  
www.illumina.com



Illumina Netherlands B.V.  
Steenoven 19  
5626 DK Eindhoven  
The Netherlands

**Sponsor australian**

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

A SE UTILIZA LA DIAGNOSTICAREA IN VITRO.

© 2024 Illumina, Inc. Toate drepturile rezervate.

**illumina**<sup>®</sup>