

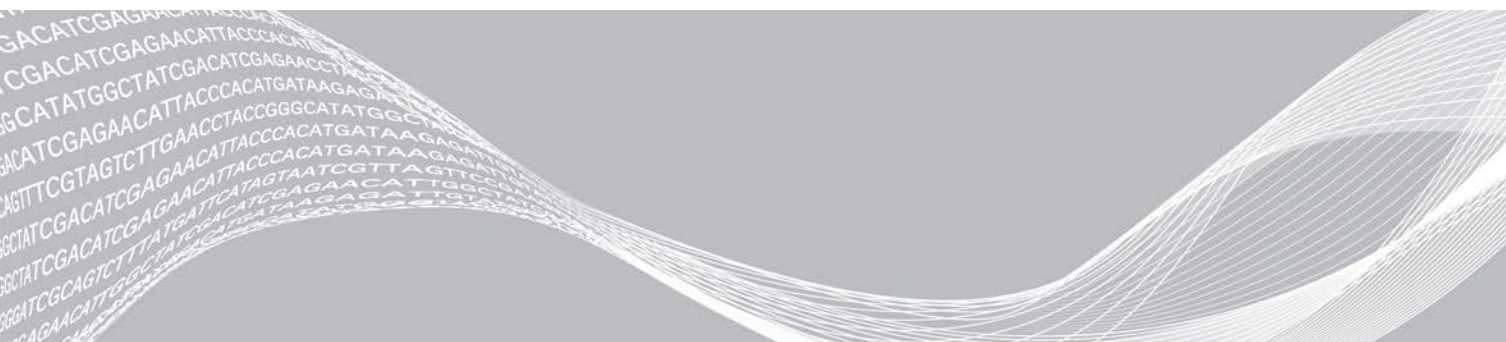
# Local Run Manager

## Modulul de analiză Variantă somatică

Ghid privind fluxul de lucru pentru NextSeq 550Dx

A SE UTILIZA LA DIAGNOSTICAREA IN VITRO

Prezentare generală	3
Introducerea informațiilor despre rulare	4
Metode de analiză	6
Vizualizarea datelor de rulare și de probă	7
Raportul de analiză	8
Fișierele de ieșire pentru analiză	9
Definirea bazelor și diversitatea indexurilor	15
Istoricul reviziilor	17
Asistență tehnică	18



Prezentul document și conținutul său constituie proprietatea Illumina, Inc. și a afiliaților 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, mărcilor sale comerciale, drepturilor sale de autor sau în temeiul dreptului comun, nicio licență și nici drepturi similare ale oricărui 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.

NECITIREA COMPLETĂ ȘI NERESPECTAREA EXPLICITĂ A TUTUROR INSTRUCȚIUNILOR CUPRINSE ÎN PREZENTUL DOCUMENT POT DUCE LA DETERIORAREA PRODUSULUI SAU PRODUSELOR, LA VĂTĂMAREA PERSOANELOR, INCLUSIV A UTILIZATORILOR SAU ALTOR PERSOANE, ȘI LA DAUNE ALE ALTOR PROPRIETĂȚI ȘI VOR 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).

© 2021 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).

## Prezentare generală

Modulul de analiză Variantă somatică al Local Run Manager este destinat utilizării împreună cu testarea cu setul de ampliconi personalizați TruSeq Dx Illumina și cu NextSeq 550Dx. Atunci când este folosită împreună cu modulul Variantă somatică, testarea este destinată pregătirii bibliotecilor folosite pentru secvențierea ADN-ului din țesut fixat în soluție de formalină și inclus în bloc de parafină (formalin-fixed, paraffin-embedded – FFPE). Testarea detectează mutațiile somatice la frecvențe scăzute ale variantelor.

Modulul de analiză evaluează regiuni scurte de ADN amplificat, sau ampliconi, pentru variante. Secvențierea focalizată a ampliconilor permite acoperirea ridicată a anumitor regiuni dintr-un număr mare de probe. Modulul de analiză efectuează analiza secundară și generarea de rapoarte din rulările de secvențiere, folosind o abordare cu catenă dublă, care implică fonduri de oligoelemente în sens direct și în sens invers. Consultați prospectul *Set de ampliconi personalizați TruSeq Dx* (nr. document 1000000029772).

Modulul de analiză Variantă somatică necesită consumabile de secvențiere pentru 300 de cicluri. Pentru mai multe informații, consultați prospectul pentru *setul de reactivi cu debit mare NextSeq 550Dx v2* sau *setul de reactivi cu debit mare NextSeq 550Dx v2.5*.

## Despre acest ghid

Acest ghid oferă instrucțiuni de configurare a parametrilor de rulare pentru secvențiere și analiză pentru modulul de analiză Variantă somatică. Pentru informații privind tabloul de bord și setările de sistem ale Local Run Manager, consultați *Ghidul de referință pentru instrumentul NextSeq 550Dx* (nr. document 1000000009513).

## Vizualizarea Local Run Manager

Interfața Local Run Manager este vizualizată în software-ul de operare NextSeq 550Dx (NOS) sau prin intermediul unui browser web. Browserul web acceptat este Chromium.



### NOTĂ

Dacă folosiți un browser neacceptat, descărcați browserul acceptat atunci când vi se solicită prin mesajul „Confirm Unsupported Browser” („Confirmare browser neacceptat”). Selectați „**here**” („aici”) pentru a descărca versiunea acceptată de Chromium.

## Vizualizarea pe monitorul instrumentului

- 1 Pentru a vizualiza interfața Local Run Manager pe monitorul instrumentului, selectați una dintre următoarele opțiuni:
  - ▶ Din Home Screen (Ecranul de întâmpinare) al NOS, selectați **Local Run Manager**. Faceți clic pe X-ul din colțul dreapta sus pentru a reveni la NOS după ce terminați.
  - ▶ Selectați pictograma Minimize NOS (Minimizare NOS), deschideți browserul web Chromium instalat pe instrument și introduceți **http://localhost** în bara de adrese. Numai utilizatorii administratori pot minimiza NOS.

## Vizualizarea de la un computer din rețea

- 1 Deschideți un browser web Chromium de pe un computer cu acces la aceeași rețea din care face parte și instrumentul și conectați-vă folosind adresa IP a instrumentului sau numele instrumentului. De exemplu, **http://myinstrument**.

## Introducerea informațiilor despre rulare

### Setarea parametrilor

- 1 Conectați-vă la Local Run Manager.
- 2 Selectați **Create Run** (Creare rulare) și selectați **Somatic Variant** (Variantă somatică).
- 3 Introduceți o denumire de rulare care identifică rularea din secvențiere prin analiză. Folosiți caractere alfanumerice, spații, caractere de subliniere sau liniuțe.
- 4 **[Optional]** Introduceți o descriere de rulare care să ajute la identificarea rulării. Folosiți caractere alfanumerice, spații, caractere de subliniere sau liniuțe.
- 5 Selectați numărul de probe și setul de indexuri din lista verticală. Luați în considerare următoarele informații atunci când faceți o selecție.
  - ▶ Lista verticală conține numere de probe cu un set de indexuri. De exemplu, 24-Set 1 indică 24 de probe care urmează a fi testate, cu indexuri din setul de indexuri 1.
  - ▶ Numerele seturilor de indexuri fac referire la seturi diferite de indexuri i5. Set 1 și Set 2 oferă ambele diversitatea indexurilor. Sunt oferite două seturi de indexuri, pentru a preveni epuizarea unui singur set.
  - ▶ Alegeți numărul de probe cel mai apropiat de numărul de probe pe care le testați. Dacă numărul exact de probe nu se află în listă, selectați numărul cel mai apropiat, dar mai mic decât numărul pe care îl testați. De exemplu, dacă doriți să testați 18 probe, selectați 16 probe.
  - ▶ Godeurile de probe și combinațiile de indexuri care îndeplinesc cerințele de diversitate a indexurilor sunt evidențiate cu verde. Dacă selectați alte godeuri și combinații de indexuri, când salvați rularea primiți o notificare dacă nu sunt îndeplinite cerințele de diversitate a indexurilor.

### Importarea de fișiere manifest pentru rulare

- 1 Asigurați-vă că manifestele pe care doriți să le importați sunt disponibile într-o locație accesibilă din rețea sau pe o unitate USB.
- 2 Selectați **Import Manifests** (Importare manifeste).
- 3 Navigați la fișierul manifest și selectați manifestele pe care doriți să le adăugați.



#### NOTĂ

Pentru a face fișierele manifest disponibile pentru toate rurile folosind modulul de analiză Somatic Variant, adăugați manifestele folosind caracteristica Module Settings (Setări modul). Această caracteristică necesită permisiuni la nivel de utilizator administrator. Pentru mai multe informații, consultați *Ghidul de referință pentru instrumentul NextSeq 550Dx (nr. document 100000009513)*.


### Specificarea probelor pentru rulare

Specificați probele pentru rulare folosind una dintre opțiunile și indicațiile care urmează.


- ▶ **Enter samples manually** (Introducere manuală a probelor) – folosiți tabelul necompletat din ecranul Create Run (Creare rulare).
- ▶ **Import samples** (Importare probe) – navigați la un fișier extern într-un format de valori separate prin virgulă (\*.csv). În ecranul Create Run (Creare rulare) este disponibil pentru descărcare un șablon.

După ce populați tabelul de probe, puteți exporta informațiile despre probe într-un fișier extern. Folosiți fișierul ca referință atunci când pregătiți bibliotecile sau importați fișierul pentru o altă rulare.

## Introducerea manuală a probelor

- 1 Introduceți o denumire de probă unică în câmpul Sample Name (Denumire probă).  
Folosiți caractere alfanumerice, liniuțe sau caractere de subliniere.  
Denumirea de probă populează automat godeul corespunzător din celălalt cumul.
- 2 **[Opțional]** Pentru probe de control pozitive sau negative, faceți clic dreapta și selectați tipul de control.  
Controlul dintr-un godeu cu probă populează automat godeul corespunzător din celălalt cumul cu același control.
- 3 **[Opțional]** Introduceți o descriere a probei în câmpul Sample Description (Descriere probă).  
Folosiți caractere alfanumerice, liniuțe sau caractere de subliniere.  
Descrierea probei populează automat godeul corespunzător din celălalt cumul.  
Descrierile de probe sunt asociate cu un ID de probă. Descrierile de probe sunt suprascrise dacă același ID de probă este folosit din nou într-o rulare ulterioară.
- 4 Selectați un adaptor Index 1 din lista verticală Index 1 (i7).  
Când folosiți godeurile cu probe sugerate, software-ul populează automat adaptoarele de indexuri i7 și i5 care îndeplinesc cerințele de diversitate a indexurilor. Dacă numărul exact al probelor pe care le testați nu se află în listă, asigurați-vă că selectați adaptoare de indexuri pentru godeuri suplimentare. Dacă trebuie să selectați indexuri pentru godeuri suplimentare sau nu folosiți combinațiile recomandate de adaptoare de indexuri, înainte de a alege indexurile, asigurați-vă că citiți *Definirea bazelor și diversitatea indexurilor la pagina 15*.
- 5 Selectați un adaptor Index 2 din lista verticală Index 2 (i5).
- 6 Selectați un fișier manifest din lista verticală Manifeste.  
Probele din Cumulul A necesită un alt manifest decât probele din Cumulul B.
- 7 Alegeți o opțiune de vizualizare, imprimare sau salvare a structurii plăcii ca referință pentru pregătirea bibliotecilor:
  - ▶ Selectați pictograma  **Print** (Imprimare) pentru a afișa structura plăcii. Selectați **Print** (Imprimare) pentru a imprima structura plăcii.
  - ▶ Selectați **Export** (Exportare) pentru a exporta informații despre probe într-un fișier extern.  
Asigurați-vă că manifestul și informațiile despre probe sunt corecte. Informațiile incorecte pot afecta rezultatele.
- 8 Selectați **Save Run** (Salvare rulare).

## Importarea probelor

- 1 Selectați **Import Samples** (Importare probe) și navigați până la locația fișierului cu informații despre probe. Există două tipuri de fișiere pe care le puteți importa.
  - ▶ Selectați **Template** (Șablon) în ecranul Create Run (Creare rulare), pentru a crea o nouă structură a plăcii. Fișierul șablon conține anteturile de coloană corecte pentru import. Introduceți informații despre probe în fiecare coloană pentru probele din rulare. Ștergeți informațiile cu titlu de exemplu din celulele nefolosite, apoi salvați fișierul.
  - ▶ Folosiți un fișier cu informații despre probe care a fost exportat din modulul de analiză Variantă somatică folosind caracteristica Export (Exportare).
- 2 Selectați pictograma **Print**  (Imprimare) pentru a afișa structura plăcii.

- 3 Selectați **Print** (Imprimare) pentru a imprima structura plăcii ca referință pentru pregătirea bibliotecilor.
- 4 **[Optional]** Selectați **Export** (Exportare) pentru a exporta informații despre probe într-un fișier extern. Asigurați-vă că manifestul și informațiile despre probe sunt corecte. Informațiile incorecte pot afecta rezultatele.
- 5 Selectați **Save Run** (Salvare rulare).

## Editarea unei rulări

Pentru instrucțiuni privind editarea informațiilor din rulare dvs. înainte de secvențiere, consultați *Ghidul de referință pentru instrumentul NextSeq 550Dx (nr. document 100000009513)*.

## Metode de analiză

Modulul de analiză Variantă somatică efectuează următoarele etape de analiză și apoi scrie fișierele de ieșire pentru analiză în folderul Alignment (Aliniere).

- ▶ Demultiplexarea citirilor de indexuri
- ▶ Generarea fișierelor FASTQ
- ▶ Alinierea la o referință
- ▶ Identificarea variantelor

## Demultiplexarea

Demultiplexarea compară fiecare secvență de citire de indexuri cu secvențele de indexuri specificate pentru rulare. Nicio valoare de calitate nu este luată în considerare în această etapă.

Citirile de indexuri sunt identificate folosind următorii pași:

- ▶ Probele sunt numerotate începând de la 1, în funcție de ordinea în care probele sunt listate pentru rulare.
- ▶ Numărul de probă 0 este rezervat pentru grupurile de celule care nu au fost alocate unei probe.
- ▶ Grupurile de celule sunt alocate unei probe atunci când secvența de indexuri se potrivește exact sau când există cel mult o singură nepotrivire per citire de indexuri.

## Generarea fișierelor FASTQ

După demultiplexare, software-ul generează fișiere de analiză intermediare în formatul FASTQ, care este un format text folosit pentru a reprezenta secvențe. Fișierele FASTQ conțin citiri pentru fiecare probă și scorurile de calitate asociate. Grupurile de celule care nu au trecut de filtrare sunt excluse.

Fiecare fișier FASTQ conține citiri pentru o singură probă și denumirea probei respective este inclusă în denumirea fișierului FASTQ. Fișierele FASTQ constituie principala introducere de date pentru aliniere. Sunt generate opt fișiere FASTQ per probă per fond de oligoelemente, patru din Citirea 1 și patru din Citirea 2, ceea ce duce la un total de 16 fișiere FASTQ per probă.

## Alinierea

În timpul etapei de aliniere, algoritmul Smith-Waterman în benzi aliniază grupurile de celule din fiecare probă la secvențele de amplicon specificate în fișierul manifest.

Algoritmul Smith-Waterman în benzi efectuează alinieri de secvențe semiglobale, pentru a determina regiunile similare dintre două secvențe. În loc să compare secvența totală, algoritmul Smith-Waterman compară segmente de toate lungimile posibile.

Fiecare citire cu perechi de baze împerecheate este evaluată în ceea ce privește alinierea sa la secvențele relevante cu rol de măsurare pentru respectiva citire.

- ▶ Citirea 1 este evaluată pe baza complementului invers al oligoelementelor cu locus specific în aval (Downstream Locus-Specific Oligos – DLSO).
- ▶ Citirea 2 este evaluată pe baza oligoelementelor cu locus specific în amonte (Upstream Locus-Specific Oligos – ULSO).
- ▶ Dacă începutul unei citiri se potrivește cu o secvență cu rol de măsurare cu nu mai mult de trei diferențe (nepotriviri sau deplasări cauzate de indelii principali), lungimea integrală a citirii este aliniată la ținta amplicon pentru respectiva secvență.
- ▶ Indelii din cadrul DLSO și ULSO nu sunt observați, dată fiind chimia de testare.

Alinierea sunt filtrate din rezultatele de aliniere pe baza ratelor de nepotrivire fie din regiunea de interes, fie din întregul amplicon, în funcție de lungimea ampliconului. Alinierea filtrate sunt scrise în fișierele de aliniere ca nealiniat și nu sunt folosite în definirea variantelor.

## Definirea variantelor



Dezvoltat de Illumina, Definitorul de variante Pisces identifică variantele prezente la frecvență scăzută în proba de ADN.

Definitorul de variante Pisces identifică SNV, MNV și indelii mici în trei etape:

- ▶ ia în considerare fiecare poziție din genomul de referință separat;
- ▶ numără bazele din poziția dată pentru citirile aliniate care suprapun poziția;
- ▶ calculează un scor de variantă care măsoară calitatea definirii folosind modelul Poisson. Variantele cu un scor de calitate sub Q30 sunt excluse.

Variantele sunt definite întâi pentru fiecare cumul separat. Apoi variantele din fiecare cumul sunt comparate și combinate într-un singur fișier de ieșire. Dacă o variantă este prezentă în ambele cumhuri și trece de toate filtrele listate în *Adnotările fișierelor VCF la pagina 12*, varianta este marcată ca PASS (REUȘITĂ) în fișierul de definire a variantelor (VCF).

## Vizualizarea datelor de rulare și de probă

- 1 Din tabloul de bord al Local Run Manager, faceți clic pe denumirea rulării.
- 2 Din fila Run Overview (Privire de ansamblu rulare), examinați metricile rulării de secvențiere.
- 3 **[Opțional]** Faceți clic pe pictograma **Copy to Clipboard** (Copiere pe clipboard)  pentru a copia calea folderului de ieșire pentru rulare.
- 4 Faceți clic pe fila Sequencing Information (Informații secvențiere) pentru a examina parametrii de rulare și informațiile despre consumabile.
- 5 Faceți clic pe fila Samples and Results (Probe și rezultate) pentru a vizualiza locația raportului de analiză.
  - ▶ Dacă analiza a fost repetată, extindeți lista verticală Select Analysis (Selectare analiză) și selectați analiza corespunzătoare.
- 6 Faceți clic pe pictograma **Copy to Clipboard** (Copiere în clipboard)  pentru a copia calea folderului Analysis (Analiză).

Pentru mai multe informații despre filele Run Overview (Privire de ansamblu rulare) și Sequencing Information (Informații secvențiere), precum și despre cum să treceți din nou analiza în lista de așteptare, consultați *Ghidul de referință pentru instrumentul NextSeq 550Dx (nr. document 1000000009513)*.

## Raportul de analiză

Rezultatele analizei sunt rezumate în fila Samples and Results (Probe și rezultate), precum și ca raport agregat în folderul Alignment (Aliniere). De asemenea, pentru fiecare probă este disponibil câte un raport în format de fișier PDF.

### Informațiile din fila Samples and Results (Probe și rezultate)

- 1 Faceți clic pe o probă din listă pentru a vedea raportul probei.

**Tabelul 1 Informații despre rulare și probe**

Antet de coloană	Descriere
Run Status (Stare rulare)	Specifică dacă rularea de secvențiere a reușit sau nu.
Total Yield (GB) (Randament total (GB))	Numărul de baze definite în rularea de secvențiere. Arată pragul de trecere și starea de reușită sau eșec.
% ≥ Q30	Procentul de citiri din rularea de secvențiere cu un scor de calitate de 30 (Q30) sau mai mare. Arată pragul de trecere și starea de reușită sau eșec.
Sample Name (Denumire probă)	Denumirea probei furnizată atunci când a fost creată rularea.
Total PF Reads (Total citiri PF)	Numărul total de citiri care trec de filtru.
Read 1% ≥ Q30 (% Citire 1 ≥ Q30)	Procentul de citiri din Citirea 1 cu un scor de calitate de 30 (Q30) sau mai mare pentru probă.
Read 2% ≥ Q30 (% Citire 2 ≥ Q30)	Procentul de citiri din Citirea 2 cu un scor de calitate de 30 (Q30) sau mai mare pentru probă.
Autosome Call Rate (Indice de apelare autozomi)	Numărul de poziții genomice din autozomi (cromozomii 1-22) care ating un prag al valorii de încredere predefinit, împărțit la numărul total de poziții genomice autozomale interogate. Indicele de apelare este descris per probă și raportat ca procent care este calculat ca 1 minus (numărul de poziții autozomale cu apelări incomplete împărțit la numărul total de poziții autozomale secvențiate).

**Tabelul 2 Informații despre raportul probei**

Antet de coloană	Descriere
Sample (Probă)	Denumirea probei furnizată atunci când a fost creată rularea.
Report Date (Dată raport)	Data la care a fost generat raportul.
Sample Information (Informații despre probă)	ID-ul probei care a fost furnizat atunci când a fost creată rularea, totalul de citiri care au trecut de filtru din probă, procentul de citiri pentru probă cu un scor de calitate de 30 (Q30) sau mai mare și indicele de apelare autozomală.
Amplicon Summary (Rezumat ampliconi)	Numărul total de regiuni de ampliconi secvențiate și lungimea totală în perechi de baze de ampliconi secvențiați din regiunile țintă, pentru proba din Cumulul A și Cumulul B, și fișierul manifest folosit pentru fiecare cumul. Fișierul manifest specifică genomul de referință și regiunile de referință vizate folosite în etapa de aliniere.
Read Level Statistics (Statistici nivel citiri)	Numărul și procentul de citiri pentru probă care acoperă fiecare poziție din referință, pentru Citirea 1 și Citirea 2 din Cumulul A și Cumulul B.
Variants Summary (Rezumat variante)	Numărul de SNV, inserții și ștergeri detectate pentru proba care a atins valorile sugerate, pentru a determina dacă rezultatele privind calitatea se încadrează într-un interval acceptabil.



Antet de coloană	Descriere
Coverage Summary (Rezumat acoperire)	Numărul total de baze aliniate împărțit la dimensiunea regiunii vizate și procentul de regiuni de ampliconi cu valori de acoperire mai mari decât pragul de acoperire scăzută de 0,2 * acoperirea medie a ampliconilor, pentru proba din Cumulul A și Cumulul B.
Coverage Plots (Digrame de acoperire)	Diagramele Acoperire după regiunea de ampliconi arată acoperirea în regiunile de ampliconi pentru probă. Regiunile cu valori de acoperire mai mici decât pragul de acoperire sunt evidențiate cu roșu. Media tuturor valorilor este indicată printr-o linie portocalie. Este furnizată câte o diagramă pentru acoperirea Cumulului A și a Cumulului B.
Software Versions (Versiuni software)	Versiunile de software atunci când proba a fost secvențiată. Include software-ul de operare NextSeq 550Dx (NOS), software-ul Local Run Manager, software-ul RTA și versiunea modulului Variantă somatică.

## Fișierele de ieșire pentru analiză

Următoarele fișiere de ieșire pentru analiză sunt generate pentru modulul de analiză Variantă somatică și furnizează rezultatele analizei pentru aliniere și definirea variantelor. Fișierele de ieșire pentru analiză se află în folderul Alignment (Aliniere).

Denumire fișier	Descriere
Demultiplexare (*.txt)	Fișiere intermediare care conțin sumarul rezultatelor demultiplexării.
FASTQ (*.fastq.gz)	Fișiere intermediare care conțin definițiile bazelor evaluate din punct de vedere al calității. Fișierele FASTQ constituie principala introducere de date pentru etapa de aliniere.
Fișiere de aliniere în format BAM (*.bam)	Conține citiri aliniate pentru o probă dată.
Fișiere de definire a variantelor per cumul în format VCF (*.vcf)	Conține variante definite în fiecare poziție, fie din cumulul în sens direct, fie din cumulul în sens invers.
Fișiere de definire a variantelor în format genom VCF (*.genome.vcf.gz)	Conține genotipul pentru fiecare poziție, fie definit ca variantă, fie definit ca referință.
Fișiere de definire a variantelor cu consens în format VCF (*.vcf.gz)	Conține variante definite în fiecare poziție din ambele cumuluri.
AmpliconCoverage_M1.tsv	Conține informații despre acoperirea per amplicon per probă pentru fiecare manifest furnizat. M# reprezintă numărul manifestului.

## Formatul fișierelor de demultiplexare

Procesul de demultiplexare citește secvența de indexuri atașată fiecărui grup de celule, pentru a determina din ce probă provine grupul de celule. Maparea dintre grupurile de celule și numărul de probe este scrisă într-un fișier de demultiplexare (\*.demux) pentru fiecare dală a Flow Cell.

Formatul de denumire a fișierului de demultiplexare este s\_1\_X.demux, unde X este numărul dalei.

Fișierele de demultiplexare încep cu un antet:

- ▶ Versiunea (număr întreg de 4 baiți), actualmente 1
- ▶ Numărul de grupuri de celule (număr întreg de 4 baiți)

Restul fișierului constă din numere de probe pentru fiecare grup de celule din dală.

Când etapa de demultiplexare este finalizată, software-ul generează un fișier de demultiplexare denumit DemultiplexSummaryF1L1.txt.

- ▶ În denumirea fișierului, **F1** reprezintă numărul Flow Cell.
- ▶ În denumirea fișierului, **L1** reprezintă numărul liniei.
- ▶ Demultiplexarea are ca rezultat un tabel cu 1 rând per dală și 1 coloană per probă, incluzând proba 0.
- ▶ Secvențele care apar cel mai frecvent în citirile de indexuri.

## Formatul fișierelor FASTQ

FASTQ este un format de fișiere bazat pe text, care conține definițiile bazelor și valorile de calitate per citire. Fiecare înregistrare conține 4 linii:

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

Identificatorul este formatat ca:

**@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber**  
**(@Instrument:IDRulare:IDFlowCell:Linie:Dală:X:Y NumCitire:SemnalizareFiltru:0:NumărProbă)**

Exemplu:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAA9#:<#<;<<<?????#=#
```

## Formatul fișierelor BAM

Un fișier BAM (\*.bam) este versiunea binară comprimată a unui fișier SAM, care este folosită pentru a reprezenta secvențe aliniate de până la 128 Mb. Formatele SAM și BAM sunt descrise în detaliu în [samtools.github.io/hts-specs/SAMv1.pdf](https://samtools.github.io/hts-specs/SAMv1.pdf).

Fișierele BAM folosesc formatul de denumire a fișierelor **SampleName\_S#.bam** (DenumireProbă\_S#.bam), unde # este numărul de probă determinat de ordinea în care probele sunt listate pentru rulare.

Fișierele BAM 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.
- ▶ **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.

Fișierele index BAM (\*.bam.bai) furnizează un index al fișierului BAM corespunzător.

## Formatul fișierelor VCF

Formatul de definire a variantelor (Variant Call Format – VCF) este un format de fișiere folosit frecvent, dezvoltat de comunitatea științifică genomică. Acesta conține informații despre variantele găsite în poziții specifice într-un genom de referință. Fișierele VCF se termină cu sufixul .vcf.

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

## Anteturile fișierelor VCF

Antet	Descriere
CHROM	Cromozomul genomului de referință. Cromozomii apar în aceeași ordine ca în fișierul de referință FASTQ.
POS	Poziția de bază unică a variantei din cromozomul de referință. Pentru SNP, această poziție este baza de referință cu varianta; pentru indeli sau ștergeri, această poziție este baza de referință imediat dinaintea variantei.
ID	Numărul rs pentru variantă obținut din dbSNP.txt, dacă este cazul. Dacă există mai multe numere rs în această locație, lista este delimitată prin punct și virgulă. Dacă nu există nicio intrare dbSNP în această poziție, este folosit un marcaj de valoare lipsă ('.') .
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 $10^{-(Q/10)}$ . De exemplu, setul de definiții Q30 are o rată de erori de 0,1%. Multe definitoare de variante atribuie scoruri de calitate pe baza modelelor lor statistice, care sunt mari în raport cu rata de erori observată.

## Adnotările fișierelor VCF

Antet	Descriere
<b>FILTER (FILTRU)</b>	<p>Dacă se trece de toate filtrele, se scrie <b>PASS (REUȘITĂ)</b> în coloana pentru filtru.</p> <ul style="list-style-type: none"> <li>• <b>LowDP</b> – aplicat siturilor cu profunzimea acoperirii sub 450x în oricare dintre cumuluri. Pentru pozițiile ampliconilor acoperite atât de citirea în sens direct, cât și de citirea în sens invers, aceasta este echivalentă cu o acoperire de citire unică de 900x.</li> <li>• <b>LowGQ</b> – calitatea genotipării (GQ) este sub limită.</li> <li>• <b>q30</b> – scorul de calitate &lt; 30.</li> <li>• <b>LowVariantFreq</b> – frecvența variantelor este mai mică decât pragul dat.</li> <li>• <b>PB</b> – decalaj cumul de măsurare. Variantă negăsită sau găsită cu frecvență scăzută în unul sau două cumuluri de măsurare.</li> <li>• <b>R3x6</b> – numărul de repetări adiacente (cu lungimea de 1 până la 3 pb) definițiilor de variante ≥ 6.</li> <li>• <b>SB</b> – decalajul catenelor este mai mare decât pragul dat.</li> </ul>
<b>INFO</b>	<p>Introducerile din coloana INFO pot include următoarele:</p> <ul style="list-style-type: none"> <li>• <b>AC</b> – numărul de alele din genotipuri pentru fiecare alelă ALT, în aceeași ordine în care sunt listate.</li> <li>• <b>AF</b> – frecvența alelică pentru fiecare alelă ALT, în aceeași ordine în care sunt listate.</li> <li>• <b>AN</b> – numărul total de alele din genotipurile definite.</li> <li>• <b>CD</b> – un marcaj care indică faptul că SNP apare în regiunea de codificare a cel puțin 1 intrare RefGene.</li> <li>• <b>DP</b> – profunzimea (numărul de definiții ale bazelor aliniate la o poziție și folosite în definirea de variante).</li> <li>• <b>Exon</b> – o listă de regiuni exonice separate prin virgulă citite din RefGene.</li> <li>• <b>FC</b> – consecință funcțională.</li> <li>• <b>GI</b> – o listă de ID-uri de gene separate prin virgulă citite din RefGene.</li> <li>• <b>QD</b> – încrederea/calitatea variantelor în funcție de profunzime.</li> <li>• <b>TI</b> – o listă de ID-uri de transcripte separate prin virgulă citite din RefGene.</li> </ul>
<b>FORMAT</b>	<p>Coloana Format enumeră câmpurile separate pe coloane. De exemplu, GT:GQ. Lista de câmpuri furnizată depinde de definatorul de variante folosit. Printre câmpurile disponibile se numără:</p> <ul style="list-style-type: none"> <li>• <b>AD</b> – intrarea în formă X,Y, unde X este numărul de definiții de referință și Y este numărul de definiții alternative.</li> <li>• <b>DP</b> – profunzimea aproximativă a citirilor; citirile cu MQ=255 sau cu perechi inadecvate sunt filtrate.</li> <li>• <b>GQ</b> – calitatea genotipului.</li> <li>• <b>GQX</b> – calitatea genotipului. GQX este valoarea minimă dintre valoarea GQ și coloana QUAL (CALITATE). În general, aceste valori sunt similare; preluarea valorii minime face din GQX măsurarea mai conservatoare a calității genotipului.</li> <li>• <b>GT</b> – genotip. 0 corespunde bazei de referință, 1 corespunde primei intrări din coloana ALT și așa mai departe. Bara oblică la dreapta (/) indică faptul că nu sunt disponibile informații despre etapizare.</li> <li>• <b>NL</b> – nivelul de zgomot; o estimare a zgomotului definiției de baze în această poziție.</li> <li>• <b>PB</b> – decalaj cumul de măsurare. Valorile mai apropiate de 0 au indicat un decalaj mai mare către un cumul de măsurare și mai puțină încredere într-o definiție de variante.</li> <li>• <b>SB</b> – decalaj catene în această poziție. Valorile negative mai mari indică un decalaj mai mic; valorile apropiate de 0 indică un decalaj mai mare.</li> <li>• <b>VF</b> – frecvența variantelor; procentul de citiri care sprijină alela alternativă.</li> </ul>
<b>SAMPLE (PROBĂ)</b>	Coloana Probă oferă valorile specificate în coloana FORMAT.

## Fișierele VCF genom

Fișierele VCF genom (gVCF) sunt fișiere VCF v4.1 care respectă un set de convenții pentru reprezentarea tuturor siturilor din genom într-un format rezonabil de compact. Fișierele gVCF (\*.genome.vcf.gz) includ toate siturile din regiunea de interes într-un singur fișier pentru fiecare probă.

Fișierul gVCF indică absența definițiilor (no-call) în pozițiile care nu trec de toate filtrele. O etichetă de genotip (GT) ./. indică absența unei definiții (no-call).

Pentru mai multe informații, consultați [sites.google.com/site/gvcftools/home/about-gvcf](https://sites.google.com/site/gvcftools/home/about-gvcf).

## Fișierele VCF per cumul și cu consens

Fluxul de lucru Variantă somatică generează 2 seturi de fișiere de definire a variantelor .

- ▶ **Per-pool VCF files (Fișiere VCF per cumul)** – conțin variante definite fie în cumulul în sens direct, fie în cumulul în sens invers. Fișierele VCF per cumul sunt scrise în folderul VariantCallingLogs.
- ▶ **Consensus VCF files (Fișiere VCF cu consens)** – conțin variante definite din ambele cumhuri. Fișierele VCF cu consens sunt scrise în folderul Alignment (Aliniere).

Fișierele VCF per cumul și cu consens includ atât fișiere VCF (\*.vcf), cât și gVCF (\*.genome.vcf), și folosesc următoarea convenție de denumire, unde S# reprezintă ordinea în care proba este listată pentru rulare:

- ▶ **Reports for all sites (Rapoarte pentru toate siturile)** – SampleName\_S#.genome.vcf
- ▶ **Reports variants only (Rapoarte doar pentru variante)** – SampleName\_S#.vcf

Software-ul compară fișierele VCF per cumul și combină datele din fiecare poziție, pentru a crea un fișier VCF cu consens pentru probă.

Definițiile de variante din fiecare cumul sunt îmbinate în fișiere VCF cu consens, folosind următoarele criterii.

Criterii	Rezultat
O definire de referință în fiecare cumul	Definire de referință
O definire de referință într-un cumul și o definire de variante în celălalt cumul	Definire de variante filtrată
Potrivirea de definiții de variante cu frecvențe similare din fiecare cumul	Definire de variante
Potrivirea de definiții de variante cu frecvențe semnificativ diferite din fiecare cumul	Definire de variante filtrată
Definiții de variante fără potriviri din fiecare cumul	Definire de variante filtrată

Metricile din fiecare cumul sunt îmbinate folosind următoarele valori.

Metrică	Valoare
Adâncime	Adăugarea de profunzimi din ambele cumhuri
Frecvența variantelor	Numărul total al variantelor împărțit la profunzimea de acoperire totală
Scorul de calitate	Valoarea minimă a ambelor cumhuri

## Fișierul de acoperire a ampliconilor

Pentru fiecare fișier manifest este generat un fișier de acoperire a ampliconilor. M# din denumirea fișierului reprezintă numărul manifestului.

Fiecare fișier include un rând de antet care conține ID-urile probelor asociate cu manifestul. Fișierul conține următoarele informații.

- ▶ ID-ul țintei, așa cum este listat în manifest.
- ▶ Profunzimea acoperirii citirilor care trec de filtru.

## Fișierele de ieșire suplimentare

Următoarele fișiere de ieșire oferă informații suplimentare sau rezumă rezultatele rulării și erorile analizei. Deși aceste fișiere nu sunt necesare pentru evaluarea rezultatelor analizei, ele pot fi folosite în scopul depanării. Toate fișierele se află în folderul Alignment (Aliniere), cu excepția cazului în care se specifică altfel.

Denumire fișier	Descriere
AnalysisLog.txt	Jurnal de procesare care descrie fiecare etapă care a avut loc în timpul analizei folderului de rulare curent. Acest fișier nu conține mesaje de eroare. Aflat în folderul Alignment (Aliniere).
AnalysisError.txt	Jurnal de procesare care enumeră orice erori care s-au produs în timpul analizei. Acest fișier va fi gol dacă nu s-a produs nicio eroare. Aflat în folderul Alignment (Aliniere).
DemultiplexSummaryF1L1#.txt	Raportează rezultatele demultiplexării într-un tabel cu 1 rând per dală și 1 coloană per probă. # reprezintă linia 1, 2, 3 sau 4 a Flow Cell. Aflat în folderul Alignment (Aliniere).
AmpliconRunStatistics.xml	Conține sumarul statisticilor specifice rulării. Aflat în folderul Alignment (Aliniere).

## Folderul de analiză

Folderul de analiză conține fișierele generate de software-ul Local Run Manager.

Relația dintre folderul de ieșire și folderul de analiză este rezumată după cum urmează:

- ▶ În timpul secvențierii, Analiza în timp real (Real-Time Analysis – RTA) populează folderul de ieșire cu fișiere generate în timpul analizei imaginilor, definirii bazelor și evaluării calității.
- ▶ RTA copiază fișierele în folderul de analiză în timp real. După ce RTA atribuie un scor de calitate fiecărei baze pentru fiecare ciclu, software-ul scrie fișierul RTAComplete.txt în ambele foldere.
- ▶ Când fișierul RTAComplete.txt este prezent, analiza începe.
- ▶ Pe măsură ce analiza continuă, Local Run Manager scrie fișierele de ieșire în folderul de analiză, apoi copiază fișierele înapoi în folderul de ieșire.





## Foldere de alinieri

De fiecare dată când analiza este trecută din nou în lista de așteptare, Local Run Manager creează un folder de alinieri denumit **Alignment\_N**, unde N este un număr secvențial.

## Structura folderelor

 **Alignment** – conține fișiere \*.bam, \*.vcf, FASTQ și fișiere specifice modulului de analiză.

 **Date and Time Stamp** – marcajul Date\_time (Dată\_oră) al analizei, în format YYYYMMDD\_HHMMSS (AAAALLZZ\_HHMMSS)

-  AnalysisError.txt
-  AnalysisLog.txt
-  aggregate.report.html
-  aggregate.report.pdf
-  aggregate.summary.csv
-  AmpliconCoverage\_M#.tsv

- 📄 AmpliconRunStatistics.xml
- 📄 Sample1.genome.vcf.gz
- 📄 Sample1.coverage.csv
- 📄 Sample1.report.pdf
- 📄 Sample1.summary.csv
- 📄 Sample1.vcf.gz
- 📄 Sample1.bam
- 📁 FASTQ
  - 📁 Sample1
    - 📄 Sample1\_L001\_R1\_001\_fastq.gz
  - 📁 Stats
    - 📄 DemuxSummaryF1L1.txt
    - 📄 FastqSummaryF1L1.txt
- 📁 Data
  - 📁 Intensities
    - 📁 BaseCalls
      - 📁 L001 – conține fișiere \*.bcl.
      - 📁 L001 – conține fișiere \*.locs.
    - 📁 RTA Logs – conține fișiere jurnal din analiza software RTA.
  - 📁 InterOp – conține fișiere binare folosite pentru a raporta metricile de rulare pentru secvențiere.
  - 📁 Logs – conține fișiere jurnal care descriu etapele efectuate în timpul secvențierii.
  - 📄 RTAComplete.txt
  - 📄 RunInfo.xml
  - 📄 RunParameters.xml

## Definirea bazelor și diversitatea indexurilor

Atunci când probele sunt secvențiate pe instrumentul NextSeq 550Dx, definirea bazelor determină o bază (A, C, G sau T) pentru fiecare grup de celule al unei date sau zonă imagistică din Flow Cell, la un ciclu specific. Instrumentul NextSeq 550Dx folosește secvențierea bicanal, care necesită doar două imagini pentru a codifica datele pentru patru baze ADN, una din canalul roșu și una din canalul verde.

Procesul pentru citirile de indexuri diferă de definirea bazelor în timpul altor citiri.

Citirile de indexuri trebuie să înceapă cu cel puțin o bază alta decât G în fiecare dintre primele două cicluri. Dacă o citire de indexuri începe cu două definiții ale bazelor pentru G, nu este generată nicio intensitate a semnalului. Semnalul trebuie să fie prezent în oricare dintre primele două cicluri pentru a asigura efectuarea demultiplexării.

La selectarea indexurilor în timpul creării rulării, apare un avertisment de diversitate scăzută dacă indexurile nu îndeplinesc cerințele de diversitate. Pentru a preveni avertismentul de diversitate scăzută, selectați secvențe de indexuri care furnizează semnal în ambele canale pentru fiecare ciclu.

- ▶ Canalul roșu – A sau C
- ▶ Canalul verde – A sau T

Acest proces de definire a bazelor asigură precizie la analizarea probelor low-plex. Pentru mai multe informații despre secvențele indexurilor dvs., consultați prospectul pentru *Set de ampliconi personalizați TruSeq Dx* (nr. document 1000000029772).

În timpul creării rulării în Local Run Manager veți alege numărul de probe care urmează a fi testate. Combinațiile de indexuri sugerate care îndeplinesc cerințele de diversitate a indexurilor sunt populate automat de software. Deși nu vi se cere să folosiți combinațiile de indexuri sugerate, acest lucru este recomandat.



## Istoricul reviziilor

Document	Data	Descrierea modificării
Nr. document 1000000030330 v04	August 2021	S-a actualizat adresa Reprezentantului autorizat în Comunitatea Europeană.
Nr. document 1000000030330 v03	Aprilie 2020	S-a actualizat adresa Reprezentantului autorizat în Comunitatea Europeană. S-a actualizat adresa sponsorului australian.
Nr. document 1000000030330 v02	Ianuarie 2019	Adăugarea de informații despre seturile de reactivi v2.5.
Nr. document 1000000030330 v01	August 2018	Marcaje de reglementare actualizate.
Nr. document 1000000030330 v00	Noiembrie 2017	Versiunea inițială.

## 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)

Numere de telefon pentru Asistență clienți Illumina

Regiune	Număr de telefon gratuit	Regional
America de Nord	+1.800.809.4566	
Australia	+1.800.775.688	
Austria	+43 800006249	+43 19286540
Belgia	+32 80077160	+32 34002973
China	400.066.5835	
Danemarca	+45 80820183	+45 89871156
Elveția	+41 565800000	+41 800200442
Finlanda	+358 800918363	+358 974790110
Franța	+33 805102193	+33 170770446
Germania	+49 8001014940	+49 8938035677
Hong Kong	800960230	
Irlanda	+353 1800936608	+353 016950506
Italia	+39 800985513	+39 236003759
Japonia	0800.111.5011	
Norvegia	+47 800 16836	+47 21939693
Noua Zeelandă	0800.451.650	
Regatul Unit	+44 8000126019	+44 2073057197
Singapore	+1.800.579.2745	
Spania	+34 911899417	+34 800300143
Suedia	+46 850619671	+46 200883979
Taiwan	00806651752	
Țările de Jos	+31 8000222493	+31 207132960
Alte țări	+44.1799.534000	

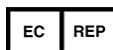
Fișe cu date de securitate (SDS) – disponibile pe site-ul web Illumina la adresa [support.illumina.com/sds.html](http://support.illumina.com/sds.html).

Documentația produselor – disponibilă pentru descărcare în format PDF de pe site-ul web Illumina. Vizitați [support.illumina.com](http://support.illumina.com), selectați un produs, apoi selectați **Documentation & Literature (Documentație și literatură)**.



Illumina  
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

CE



Illumina Netherlands B.V.  
Steenoven 19  
5626 DK Eindhoven  
Țările de Jos

**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**

© 2021 Illumina, Inc. Toate drepturile rezervate.

**illumina®**