

DRAGEN for Illumina DNA Prep with Enrichment Dx pe NextSeq 550Dx

Ghidul de utilizare a aplicației

PROPRIETATEA ILLUMINA

Document nr. 200025238 v00

Februarie 2023

A SE UTILIZA LA DIAGNOSTICAREA IN VITRO.

Prezentul document și conținutul său 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 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.

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).

© 2023 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.

Istoricul versiunilor

Document	Data	Descrierea modificării
200025238 v00	Februarie 2023	Versiunea inițială.

Cuprins

Istoricul versiunilor	iii
Prezentare generală	1
Metode de analiză	1
Crearea unei rulări planificate	5
Setări	7
Fișier manifest	8
Filtrarea zgomotului (opțional)	9
Rezultatele analizei	9
Fișiere FASTQ	10
Fișiere BAM	11
Fișiere VCF	11
Retrimiteri analiză în coadă	18
Asistență tehnică	19

Prezentare generală

Aplicația DRAGEN for Illumina DNA Prep with Enrichment Dx (DRAGEN for IDPE Dx) este utilizată pentru a planifica și a efectua analiza secundară a bibliotecilor IDPE Dx generate pentru secvențiere pe NextSeq 550Dx.

DRAGEN for IDPE Dx acceptă secvențierea pentru analiză atunci când se utilizează împreună cu Pregătirea bibliotecii Illumina DNA Prep with Enrichment Dx, NextSeq 550Dx și Illumina DRAGEN Server for NextSeq 550Dx.

Metode de analiză

DRAGEN for IDPE Dx efectuează demultiplexarea, generarea FASTQ, maparea citirii, alinierea la un genom de referință și definirea variantelor mici în funcție de fluxurile de lucru selectate:

- Generare FASTQ
- Generare Germline FASTQ și VCF
- Generare Somatic FASTQ și VCF

NOTĂ Compresia ORA este disponibilă pentru utilizare cu toate cele trei fluxuri de lucru. DRAGEN ORA Compression este un software de compresie complet fără pierderi care creează un fișier cu o extensie de Original Read Archive (*.ora). Formatul ora este un format de compresie bazat pe referință pentru fișierele FASTQ și este conceput pentru compresie/decompresie foarte rapidă și raport de compresie ridicat.

Generare FASTQ

Secvențele asamblate sunt scrise în fișiere FASTQ per probă. Fișierele FASTQ sunt fișiere text care conțin date de secvențiere și scoruri de calitate pentru o singură probă. Pentru fiecare probă sunt generate fișiere FASTQ separate pentru fiecare bandă de celule de flux, pentru fiecare citire de secvențiere. Denumirea probei, așa cum este specificată în timpul configurării rulării, este inclusă în numele fișierului FASTQ. Fișierele FASTQ reprezintă principala introducere de date pentru aliniere. Primul pas al generării FASTQ este demultiplexarea. Demultiplexarea atribuie grupuri de celule care trec filtrul la o probă prin compararea fiecărei secvențe de citire de indexuri cu secvențele index specificate pentru rulare. În această etapă, nu este luată în considerare nicio valoare calitativă. Citirile de indexuri sunt identificate folosind următorii pași:

- Probele sunt numerotate începând de la 1, în funcție de ordinea în care sunt enumerate 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.

Software-ul include compresie ORA pentru compresia fișierelor FASTQ. Acest format poate fi activat opțional. La utilizarea formatului ORA (*.ora), suma de control md5 a conținutului FASTQ este păstrată după un ciclu de compresie și decompresie pentru a asigura o compresie fără pierderi.

Maparea și alinierea ADN-ului

După generarea FASTQ, citirile sunt mapate și aliniate la un genom de referință. Prima etapă a mapării este generarea de selecții din citire, căutând apoi potriviri exacte în genomul de referință. Aceste rezultate sunt rafinate apoi rulând alinieri Smith-Waterman complete în locațiile cu cea mai mare densitate a potrivirilor de selecții. Acest algoritm bine documentat funcționează comparând fiecare poziție a citirii cu toate pozițiile candidate ale referinței. Aceste comparații corespund unei matrice de alinieri potențiale între citire și referință. Pentru fiecare dintre aceste poziții de aliniere a candidaților, Smith-Waterman generează scoruri care sunt utilizate pentru a evalua dacă cea mai bună aliniere care trece prin acea celulă matrice ajunge la aceasta printr-o potrivire sau nepotrivire (mișcare diagonală), o deleție (mișcare orizontală) sau o inserție (mișcare verticală) de nucleotide. O potrivire între citire și referință oferă un bonus pentru scor, iar o nepotrivire sau indel impune o penalizare. Calea cu cel mai mare punctaj general la trecerea prin matrice este alinierea aleasă. Algoritmul este accelerat de hardware pe cardurile cu blocuri de circuite logice programabile în câmp (FPGA) DRAGEN. Genomul de referință utilizat în aplicație este creat din FASTA UCSC hg19 cu opțiunea DRAGEN pentru a crea un tabel hash sensibil la alternative bazat pe preluarea caracteristicilor genomului.

Definirea variantelor DRAGEN Germline

Definitorul de variante mici DRAGEN Germline preia citirile de ADN mapate și aliniate ca date de intrare și numește polimorfismele uninucleotidice (SNP-urile) și inserția sau deleția (indeli) printr-o combinație de detectare la nivel de coloană și asamblare *de novo* locală de haplotipuri. Pentru a activa Definitorul de variante mici DRAGEN Germline, selectați fluxul de lucru dedicat variantelor de linie germinală.

Definirea variantelor de linie germinală este utilizată în mod obișnuit pentru probele de la nivelul liniei germinale, unde se cunoaște că ploidia este doi. Regiunile de referință care pot fi definite sunt identificate mai întâi cu o acoperire suficientă a alinierii. În cadrul acestor regiuni de referință, o scanare rapidă a citirilor sortate identifică regiunile active, care sunt centrate pe coloanele de acumulare cu dovezi ale unei variante. Regiunile active sunt completate cu un context suficient pentru a acoperi conținutul semnificativ, care nu este de referință din apropiere. Dacă există dovezi de indeli, regiunile active primesc acoperire suplimentară.

Citirile aliniate sunt decupate din cadrul fiecărei regiuni active și asamblate într-un grafic De Bruijn. Marginile citirilor decupate sunt ponderate în funcție de numărul de observații, cu secvența de referință drept structură fundamentală. După o anumită curățare și simplificare a graficului, toate căile de la sursă la scufundare sunt extrase ca haplotipuri candidate. Fiecare haplotip este aliniat cu ajutorul algoritmului Smith-Waterman la genomul de referință pentru a identifica variantele pe care le reprezintă. Acest set de evenimente poate fi augmentat printr-o detectare bazată pe poziție. Pentru fiecare pereche de haplotipuri citite, probabilitatea $P(r|H)$ de a observa citirea, presupunând că haplotipul este proba de pornire reală, este estimată utilizând un model Markov ascuns (HMM) pereche.

Scanând după poziția de referință la nivelul regiunii active, genotipurile candidate se formează din combinații diploide de evenimente ale variantelor (SNP-uri sau indeli). Pentru fiecare eveniment (incluzând referința), probabilitatea condiționată $P(r|e)$ de a observa fiecare citire suprapusă este estimată ca valoarea maximă $P(r|H)$ pentru haplotipurile care susțin evenimentul. Acestea sunt combinate în probabilitatea condiționată $P(r|e1e2)$ pentru un genotip (pereche de evenimente) și înmulțite pentru a genera probabilitatea condiționată $P(R|e1e2)$ de a observa întreaga acumulare de citiri. Utilizând formula Bayes, se calculează probabilitatea posterioară $P(e1e2|R)$ a fiecărui genotip diploid, definind astfel câștigătorul.

DRAGEN for IDPE Dx aplică filtrarea automată. Consultați [Adnotările fișierului VCF pentru fluxul de lucru Germline la pagina 13](#) pentru mai multe informații.

Definirea variantelor DRAGEN Somatic

Definitorul de variante mici DRAGEN Somatic preia citirile de ADN mapate și aliniate ca date de intrare și numește SNV-urile și indeli prin asamblarea *de novo* locală de haplotipuri într-o regiune activă. Pentru a activa Definitorul de variante mici DRAGEN Somatic, selectați o aplicație pentru variante somatice.

Definirea variantelor somatice este utilizată de obicei pentru probele tumorale. Cu acest flux de lucru, DRAGEN nu face nicio presupunere cu privire la ploidie, ceea ce permite detectarea alelelor de frecvență joasă. Pentru locii cu acoperire de până la 100x în proba tumorală, DRAGEN are un prag de detecție la frecvențe alelice ale variantelor de 5 %. Limita se scalează odată cu creșterea adâncimii pentru fiecare locus și se jumătățește de fiecare dată când acoperirea se dublează peste 100x. Regiunile de referință care pot fi definite sunt identificate mai întâi cu o acoperire suficientă a alinierii. În cadrul acestor regiuni de referință, o scanare a citirilor sortate identifică regiunile active, care sunt centrate pe coloanele de acumulare cu dovezi ale unei variante la citirile tumorale. Regiunile active sunt completate cu un context suficient pentru a acoperi conținutul semnificativ, care nu este de referință din apropiere. Dacă există dovezi de indeli, regiunile active primesc acoperire suplimentară.

Citirile aliniate sunt decupate din cadrul fiecărei regiuni active și asamblate într-un grafic De Bruijn. Marginile citirilor decupate sunt ponderate în funcție de numărul de observații, cu secvența de referință drept structură fundamentală. După o anumită curățare și simplificare a graficului, toate căile de la sursă la scufundare sunt extrase ca haplotipuri candidate. Fiecare haplotip este aliniat cu ajutorul algoritmului Smith-Waterman la genomul de referință pentru a identifica variantele pe care le reprezintă. Pentru fiecare pereche de haplotipuri citite, probabilitatea $P(r|H)$ de a observa citirea este estimată utilizând un model Markov ascuns (HMM) pereche, presupunând că haplotipul este proba de pornire reală.

Pentru a determina scorul pentru limita de detecție a tumorii (TLOD), Definitorul de variante mici DRAGEN Somatic scanează mai întâi după poziția de referință pentru fiecare eveniment somatic candidat, precum și evenimentul de referință la nivelul regiunii active. Probabilitatea condiționată $P(r|e)$ de a observa fiecare citire suprapusă este estimată ca valoarea maximă $P(r|H)$ pentru haplotipurile care susțin evenimentul. Acestea sunt combinate în probabilitatea condiționată $P(r|E)$ pentru o ipoteză de eveniment, E, implicând un amestec de alele somatice de referință și candidat pe o gamă de frecvențe posibile ale alelelor și înmulțite pentru a genera probabilitatea condiționată $P(R|E)$ de a observa întreaga acumulare de citiri. De acolo, se calculează un scor TLOD ca dovadă a faptului că o alelă ALT este prezentă în proba tumorală într-un anumit locus.

DRAGEN for IDPE Dx aplică filtrarea automată. Consultați [Adnotările fișierului VCF pentru fluxul de lucru Somatic la pagina 16](#) pentru mai multe informații.

Crearea unei rulări planificate

Utilizați pașii următori pentru a configura o rulare în Illumina Run Manager fie pe NextSeq 550Dx, fie utilizând un browser pe un computer din rețea. Utilizați un browser pe un computer conectat la rețea dacă doriți să importați date despre probe. Consultați Ghid software pentru Illumina Run Manager for NextSeq 550Dx (document nr. 200025239) pentru instrucțiuni privind accesarea Illumina Run Manager de pe un computer conectat la rețea.

Există două moduri diferite de a crea o nouă rulare planificată:

- **Import Run** (Importare rulare) — Utilizați o fișă de probe de la o rulare existentă ca șablon pentru o rulare nouă. Consultați Ghid software pentru Illumina Run Manager for NextSeq 550Dx (document nr. 200025239) pentru informații despre modul de importare a unei rulări.
- **Create Run** (Creare rulare) — Introduceți manual parametrii de rulare. Instrucțiunile următoare descriu modul în care se creează o rulare.

NOTĂ Câmpurile de introducere obligatorii din interfața cu utilizatorul sunt marcate cu un simbol reprezentând un asterisc (*).

Aplicație

1. Din fila Planned (Planificate) a ecranului Runs (Rulări), selectați **Create Run** (Creare rulare).
2. Selectați aplicația DRAGEN for Illumina DNA Prep with Enrichment Dx, apoi selectați **Next** (Înainte).

Setările rulării

1. Pe ecranul Run Settings (Setări rulare), introduceți un nume unic de rulare. Numele rulării identifică rularea din secvențiere prin analiză.
2. **[Opțional]** Introduceți o descriere de rulare care să ajute la identificarea ulterioară a rulării.
3. Selectați setul/seturile de adaptoare de indexare utilizat(e) în timpul pregătirii bibliotecii.
4. Examinați Lungimea de citire și modificați-o dacă este necesar. Citirea 1 și Citirea 2 au o valoare implicită de 151 de cicluri. Indexul 1 și Indexul 2 au o valoare fixă de 10 cicluri și nu se pot modifica.
5. **[Opțional]** Introduceți ID-ul unei eprubete din bibliotecă.
6. Selectați **Next** (Înainte).

Datele probei

Datele probei includ ID-ul probei, Poziția godeului (poziția godeului pe placa de indexare) și Denumirea bibliotecii. Când se utilizează indexul A&B, Poziția godeului include și identificatorul Plăcii.

Există două modalități de introducere a datelor probei:

- **Import Samples** (Importare probe) — Utilizați un fișier șablon disponibil pentru descărcare pe ecranul Sample Data (Date probă).
- **Manual** — Introduceți datele probei direct în tabelul de pe ecranul Sample Data (Date probă).

Importarea probelor

Atunci când se planifică o rulare de secvențiere folosind un browser pe un computer conectat la rețea, un fișier șablon (*.csv) este disponibil pentru descărcare pe ecranul Sample Data (Date probă). Fișierul șablon nu este disponibil pentru descărcare la accesarea Illumina Run Manager prin intermediul software-ului sistemului de operare NextSeq 550Dx. Pentru a introduce date despre probe utilizând funcția Import Samples (Importare probe), parcurgeți următorii pași.

NOTĂ Finalizați pașii din Run Settings (Setări rulare) înainte de a continua.

1. Selectați **Download Template** (Descărcare șablon) pentru a descărca un fișier CSV gol.
2. Din fișierul șablon, introduceți datele probei, apoi salvați fișierul. Denumirea bibliotecii este opțională.

NOTĂ Atunci când se utilizează Indexul A&B, datele pentru coloana B trebuie să includă atât poziția plăcii, cât și poziția godeului (poziția godeului pe placa de indexare). Exemplu: A-A01, A-A02, A-A03.

3. Selectați **Import Samples** (Importare probe) și navigați la fișierul șablon care conține informațiile despre datele probei de la pasul anterior.
4. Selectați **Open** (Deschidere), **Proceed** (Continuare) și apoi **Next** (Înainte).

NOTĂ Modificarea ID-ului probei înainte de a selecta Next (Înainte) poate duce la o eroare. Finalizați configurarea rulării înainte de a efectua modificări pentru a evita erorile.

Introducerea manuală a probelor

Utilizați tabelul de pe ecranul Sample Data (Date probă) pentru a introduce manual datele probei.

1. Introduceți un ID de probă unic în câmpul Sample ID (ID probă).
2. Utilizați **Well Position** (Poziție godeu) (Indexul A sau Indexul B) sau **Plate - Well Position** (Placă - Poziție godeu) (Indexul A&B) pentru a selecta indexul asociat pentru probe. Câmpurile Index i7, Index 1, Index i5 și Index 2 se populează în mod automat.
3. **[Opțional]** Introduceți o denumire de bibliotecă.
4. Adăugați rânduri și repetați pașii 1–3 după cum este necesar, până când toate probele au fost adăugate în tabel. Puteți adăuga mai multe rânduri simultan introducând mai întâi numărul de rânduri ce vor fi adăugate și apoi selectând pictograma +. De asemenea, puteți elimina rânduri selectând caseta de lângă numărul rândului, apoi făcând clic pe pictograma coș de gunoi.
5. Selectați **Next** (Înainte).

Setări analiză

1. Selectați fluxul de lucru pentru analiză dorit:
 - Generare FASTQ
 - Generarea FASTQ și VCF pentru un flux de lucru de tip linie de germeni (este necesar un fișier manifest)
 - Generarea FASTQ și VCF pentru un flux de lucru somatic (este necesar un fișier manifest)
2. **[Opțional]** Opțiunea **Generate ORA compressed FASTQs** (Generare FASTQ-uri comprimate ORA) este activată în mod implicit. Compresia ORA FASTQ comprimă fără pierderi fișierele FASTQ de până la 5x în comparație cu fastq.gz. Debifați **Generate ORA compressed FASTQs** (Generare FASTQ-uri comprimate ORA) dacă se preferă date necomprimate (fastq.gz).
3. Pentru fluxurile de lucru Germline și Somatic, este necesar un fișier manifest. Utilizați meniul derulant **Manifest File Selection** (Selectare fișier manifest) pentru a selecta un fișier manifest. Manifestul este un fișier BED delimitat cu tabulatori (* .bed), care specifică numele și locațiile regiunilor de referință vizate. Pentru mai multe informații, consultați [Fișier manifest la pagina 8](#).
4. **[Opțional]** Pentru fluxurile de lucru Somatic, utilizați meniul derulant **Noise File Selection** (Selectare fișier zgomot) pentru a selecta un fișier de zgomot sistematic. Un fișier BED (* .bed .gz) cu nivel de zgomot specific unității poate fi specificat pentru filtrarea zgomotului sistematic. Pentru mai multe informații, consultați [Filtrarea zgomotului \(opțional\) la pagina 9](#).
5. Selectați **Next** (Înainte).

Rulare revizuire

1. Pe ecranul Review (Revizuire), examinați informațiile pentru Run Settings (Setări rulare), Sample Data (Date probă) și Analysis Settings (Setări analiză).
2. Selectați **Save** (Salvare).
Rularea este salvată în fila Planned (Planificate) de pe ecranul Runs (Rulări).

Setări

Pentru a vizualiza sau a modifica setările aplicației DRAGEN for IDPE Dx, selectați mai întâi pictograma Applications (Aplicații) din ecranul principal. Apoi selectați aplicația pe care doriți să o vizualizați sau să o modificați. Este necesar un cont de Administrator pentru a modifica setările.

Configurație

Ecranul Configuration (Configurație) afișează următoarele setări ale aplicației:

- **Library Prep Kits** (Seturi de pregătire a bibliotecii) – afișează setul de pregătire a bibliotecii implicit pentru aplicație. Această setare nu poate fi modificată.
- **Index Adapter Kits** (Seturi de adaptoare de indexare) – afișează setul de adaptoare de indexare implicit pentru aplicație. Această setare nu poate fi modificată.
- **Read lengths** (Lungimi citire) – lungimile citirilor sunt setate la 151 pentru aplicație în mod implicit, însă pot fi modificate în timpul creării rulării.

- **Manifest and Noise Files** (Fișiere manifest și de zgomot) – încărcați și modificați setările pentru fișierele manifest și de zgomot.
 - Selectați **Upload File** (Încărcare fișier) pentru a încărca fișierele ce vor fi utilizate în analiză.
 - Selectați butonul radio **Default** (Implicit) pentru a seta fișierul ca manifest sau fișier de zgomot implicit selectat în timpul creării rulării atunci când este selectată aplicația.
 - Bifați caseta de selectare **Enabled** (Activat) pentru a seta fișierul ce se va afișa în meniul derulant în timpul creării rulării.

Permișiuni

Utilizați casetele de selectare de pe ecranul Permissions (Permișiuni) pentru a gestiona accesul utilizatorilor pentru aplicație.

Fișier manifest

Când utilizați DRAGEN for IDPE Dx, este necesară introducerea unui fișier manifest pentru următoarele fluxuri de lucru:

- Generarea FASTQ și VCF pentru un flux de lucru de tip linie de germeni
- Generarea FASTQ și VCF pentru un flux de lucru somatic

Fișierul manifest este un fișier de text delimitat cu tabulatori care folosește formatul BED (*.bed) și care specifică numele și locațiile regiunilor de referință vizate. Secțiunea principală a fișierului manifest este secțiunea Regions (Regiuni), care trebuie să conțină următoarele coloane de date:

Coloană	Descriere
Denumire	Nume unic specificat de utilizator pentru țintă
Cromozom	Locația cromozomului (de ex., chr10, chr5 etc.)
Start	Indicele bazat pe 1 pentru poziția inițială a țintei
Oprire	Indicele bazat pe 1 pentru poziția finală a țintei
Lungime sondă în amonte	Lungimea sondei din amonte. Pentru aplicația DRAGEN for IDPE Dx, aceasta trebuie setată la 0.
Lungime sondă în aval	Lungimea sondei din aval. Pentru aplicația DRAGEN for IDPE Dx, aceasta trebuie setată la 0.

NOTĂ Pentru analiză este necesar un format valid de fișier manifest. DRAGEN va opri analiza dacă fișierul manifest este nevalid.

Filtrarea zgomotului (opțional)

Filtrarea zgomotului sistematic este disponibilă pentru definirea variantelor somatice, putând fi utilizată pentru a reduce definițiile fals pozitive, luând în calcul zgomotul specific locației. Fișierul de zgomot sistematic este generat mai întâi prin colectarea a aproximativ 50 de probe normale (preferabil specifice panoului, pregătirii bibliotecii și sistemului de secvențiere), iar apoi suma frecvențelor alelice sub 30 % la fiecare locație cu acoperire suficientă este împărțită la numărul total de probe (frecvențele alelice peste 30 % sunt presupuse a fi variante ale liniei germinale și nu zgomot). Odată ce sunt generate valorile de zgomot, variantele somatice detectate în acea locație vor fi filtrate.

Filtrul poate fi utilizat în modul Tumor-Normal (Probă tumorală - normală), dar este util în special pentru rulările Tumor-Only (exclusiv probe tumorale) unde nu este disponibilă o probă normală corespondentă. Fișierul de zgomot sistematic trebuie să utilizeze un fișier BED care are o extensie de fișier (*.bed.gz) și trebuie să includă patru coloane: Niveluri de zgomot pentru Cromozom, Inițial, Final și specific locației pentru fiecare rând. Filtrarea zgomotului sistematic este opțională.

Rezultatele analizei

Rulările în curs în prezent sunt afișate în fila Active. Rulările finalizate sunt afișate în fila Completed (Finalizate). DRAGEN for IDPE Dx creează un folder de analiză cu denumire unică pentru fiecare analiză, care este separat de folderul care conține datele de secvențiere. Folderul de analiză include următoarele informații:

- Fișier manifest utilizat
- Versiune software
- ID-uri probe
- Total citiri aliniat
- Procent de citiri aliniat per probă
- Numărul de SNV-uri definite per probă
- Numărul de indeli definiți per probă
- Statistici privind acoperirea

Fișierele de ieșire ale analizei

Locația folderului de analiză este specificată de setarea External Storage for Analysis (Spațiu extern de stocare pentru analiză). Consultați Ghid software pentru Illumina Run Manager for NextSeq 550Dx (document nr. 200025239) pentru mai multe informații despre setarea External Storage for Analysis (Spațiu extern de stocare pentru analiză).

Pe ecranul Run details (Detaliile rulării), câmpul External Location (Locație externă) oferă calea pentru datele de secvențiere. Denumirea unică a folderului de analiză este furnizată în câmpul Analysis Output Folder (Folder de ieșire pentru analiză) de pe ecranul Run details (Detaliile rulării). Fișierele exacte generate depind de fluxul de lucru al analizei utilizat. Următoarele fișiere de ieșire pentru analiză sunt generate de aplicație.

NOTĂ Dacă intervine o eroare de limitare a lungimii maxime a căii fișierului la accesarea fișierelor de ieșire pentru analiză, încercați să mutați fișierul într-o locație mai scurtă sau utilizați o altă metodă pentru a deschide fișierul.

Fișier de ieșire	Descriere
Raport de sinteză privind variantele (*.pdf)	Conține un rezumat al informațiilor fișierului, versiuni de software, informații despre probă, statisticile nivelului de citire și SNV, rezumate privind inserțiile, delețiile și acoperirea. Numai fluxurile de lucru de linie germinală și somatice generează un raport privind variantele.
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. Când este selectată compresia ORA, se utilizează extensia de fișier *.fastq.ora.
Fișiere BAM de aliniere (*.bam)	Conține citiri aliniat pentru o probă dată.
Fișiere VCF genom (*.gvcf.gz)	Conține genotipul pentru fiecare poziție, fie definit ca variantă, fie definit ca referință.
Fișiere VCF (*.vcf.gz)	Conține variante definite în fiecare poziție.
Raportul parametrilor de rulare (*.csv)	Conține parametrii de calitate pentru rulare, inclusiv randamentul total neindexat și scorul Q30.

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
- Scorurile de calitate Phred într-un format codificat ASCII + 33

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=AAAAAAAAA9#:<#<;<<<????#=#
```

Fișiere BAM

Un fișier BAM (*.bam) este versiunea binară comprimată a unui fișier SAM (harta alinierii secvenței), care este folosită pentru a reprezenta secvențe aliniate de până la 128 Mb. Fișierele BAM folosesc formatul de denumire a fișierelor `DenumireProbă_S#.bam`, unde # este numărul de probă determinat de ordinea în care probele sunt enumerate pentru rulare. În modul multinod, S# este setat la S1, indiferent de ordinea probei.

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.
- 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.

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

Fișiere VCF

Fișierele în formatul de definire a variantelor (*.vcf) conțin informații despre variantele situate în poziții specifice într-un genom de referință.

Antetul fișierelor VCF include versiunea de format a fișierului VCF, 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 singură variantă.

Tabelul 1 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.

Antet	Descriere
ID	Numărul rs (SNP de referință) pentru SNP obținut din <code>dbSNP.txt</code> , dacă este cazul. Dacă există mai multe numere rs în această locație, lista este delimitată prin punct și virgulă. Dacă nu există o intrare dbSNP în această poziție, se folosește 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.

Tabelul 2 Adnotările fișierului VCF pentru fluxul de lucru Germline

Antet	Descriere
FILTER (FILTRU)	<p>Dacă se trece de toate filtrele, apare consemnat PASS (REUȘITĂ) în coloana pentru Filter (Filtru). Intrările posibile pentru FILTER (FILTRU) includ:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL – se aplică dacă scorul QUAL pentru varianta SNP nu atinge pragul • DRAGENIndelHardQUAL – se aplică dacă scorul QUAL pentru varianta indel nu atinge pragul • LowDepth – locație eliminată prin filtrare deoarece profunzimea de acoperire nu atinge pragul • LowGQ – locație eliminată prin filtrare deoarece calitatea genotipului nu atinge pragul • PloidyConflict – definirea genotipului de la definatorul de variante nu este în concordanță cu ploidia cromozomilor • base_quality – locație eliminată prin filtrare deoarece calitatea mediană a bazei pentru citirile ALT în acest locus nu atinge pragul • filtered_reads – locație eliminată prin filtrare deoarece o fracțiune prea mare de citiri a fost eliminată prin filtrare • fragment_length – 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 • low_depth – locație eliminată prin filtrare deoarece adâncimea de citire este prea scăzută • low_frac_info_reads – locație eliminată prin filtrare deoarece fracțiunea de citiri informative este sub prag • low_normal_depth – locație eliminată prin filtrare deoarece adâncimea de citire a probei normale este prea scăzută • long_indel – locație eliminată prin filtrare deoarece lungimea indelului este prea mare • mapping_quality – locație eliminată prin filtrare deoarece calitatea mediană a mapării pentru citirile ALT în acest locus nu atinge pragul • multiallelic – locație eliminată prin filtrare deoarece mai mult de două alele ALT trec de LOD tumorală • non_homref_normal – locație eliminată prin filtrare deoarece genotipul probei normale nu este o referință homozigotă • no_reliable_supporting_read – locație eliminată prin filtrare deoarece nu există o citire somatică de susținere fiabilă • panel_of_normals – văzut în cel puțin o probă din panoul de vcf cu probe normale • read_position – locație eliminată prin filtrare deoarece mediana distanțelor dintre începutul/sfârșitul citirii și acest locus este sub prag • RMxNRepeatRegion – locație eliminată prin filtrare deoarece întreaga sau o parte a alelei variantei este o repetare a referinței • strand_artifact – locație eliminată prin filtrare din cauza unui decalaj sever al catenelor • str_contraction – locație eliminată prin filtrare din cauza erorii PCR suspectate, unde alela ALT este cu o unitate de repetare mai mică decât referința • too_few_supporting_reads – locație eliminată prin filtrare deoarece există prea puține citiri de susținere în proba tumorală • weak_evidence – scorul variantei somatice nu atinge pragul

Antet	Descriere
INFO (INFORMAȚII)	<p>Intrările posibile pentru INFO (Informații) includ:</p> <ul style="list-style-type: none"> • AC – numărul de alele din genotipuri pentru fiecare alelă ALT, în aceeași ordine în care sunt enumerate. • AF – frecvența alelică pentru fiecare alelă ALT, în aceeași ordine în care sunt enumerate. • AN – numărul total de alele din genotipurile definite. • DB – calitatea de membru dbSNP. • FS – valoare p pe scara Phred, utilizând testul exact al lui Fisher pentru a detecta decalajul catenelor. • QD – încrederea/calitatea variantelor în funcție de profunzime. • R2_5P_bias – scor bazat pe decalajul dintre corespondenți și distanța de la capătul 5 al catenei. • SOR – raportul probabilităților simetrice a tabelului de contingență 2x2 pentru detectarea decalajului dintre catene. • DP – profunzimea aproximativă a citirilor (informative și neinformative); este posibil ca unele citiri să fi fost eliminate prin filtrare în baza mapq etc. • END – poziție de final a intervalului. • FractionInformativeReads – fracția de citiri informative din totalul citirilor. • MQ – calitatea mapării RMS. • MQRankSum – scorul Z la testul sumei rangurilor Wilcoxon pentru calitățile de mapare Alt vs. Ref. • ReadPosRankSum – scorul Z la testul sumei rangurilor Wilcoxon pentru eroarea sistematică de poziție Alt vs. Ref. • SOMATIC – cel puțin o variantă în această poziție este somatică.

Antet	Descriere
FORMAT	<p>Coloana Format enumeră câmpuri separate de două puncte: De exemplu, GT:GQ. Printre câmpurile disponibile se numără:</p> <ul style="list-style-type: none"> • AD – profunzimi alelice (numărând doar citirile informative din totalul citirilor) pentru alelele de REF și ALT în ordinea enumerată. • AF – fracții de alele pentru alelele ALT în ordinea enumerată. • DP – profunzimea aproximativă a citirilor (sunt filtrate citirile cu MQ=255 sau cu perechi inadecvate). • F1R2 – numărul de citiri în orientarea perechii F1R2 care susține fiecare alelă. • F2R1 – numărul de citiri în orientarea perechii F2R1 care susține fiecare alelă. • GT – 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. • MB – statisticile componentelor pentru fiecare probă pentru a detecta decalajul dintre corespondenți. • PS – 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. • SB – statisticile componentelor pentru fiecare probă care cuprind testul exact al lui Fisher pentru a detecta decalajul catenelor. • SQ – calitatea somatică.
SAMPLE (PROBĂ)	Coloana Sample (Probă) prezintă valorile specificate în coloana FORMAT.

Tabelul 3 Adnotările fișierului VCF pentru fluxul de lucru Somatic

Antet	Descriere
FILTER (FILTRU)	<p>Dacă se trece de toate filtrele, apare consemnat PASS (REUȘITĂ) în coloana pentru Filter (Filtru). Intrările posibile pentru FILTER (FILTRU) includ:</p> <ul style="list-style-type: none"> • base_quality – locație eliminată prin filtrare deoarece calitatea mediană a bazei pentru citirile ALT în acest locus nu atinge pragul • filtered_reads – locație eliminată prin filtrare deoarece o fracțiune prea mare de citiri a fost eliminată prin filtrare • fragment_length – 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 • low_depth – locație eliminată prin filtrare deoarece adâncimea de citire este prea scăzută • low_frac_info_reads – locație eliminată prin filtrare deoarece fracțiunea de citiri informative este sub prag • low_normal_depth – locație eliminată prin filtrare deoarece adâncimea de citire a probei normale este prea scăzută • long_indel – locație eliminată prin filtrare deoarece lungimea indelului este prea mare • mapping_quality – locație eliminată prin filtrare deoarece calitatea mediană a mapei pentru citirile ALT în acest locus nu atinge pragul • multialelic – locație eliminată prin filtrare deoarece mai mult de două alele ALT trec de LOD tumorală • non_homref_normal – locație eliminată prin filtrare deoarece genotipul probei normale nu este o referință homozigotă • no_reliable_supporting_read – locație eliminată prin filtrare deoarece nu există o citire somatică de susținere fiabilă • panel_of_normals – văzut în cel puțin o probă din panoul de vcf cu probe normale • read_position – locație eliminată prin filtrare deoarece mediana distanțelor dintre începutul/sfârșitul citirii și acest locus este sub prag • RMxNRepeatRegion – locație eliminată prin filtrare deoarece întreaga sau o parte a alelei variantei este o repetare a referinței • strand_artifact – locație eliminată prin filtrare din cauza unui decalaj sever al catenelor • str_contraction – locație eliminată prin filtrare din cauza erorii PCR suspectate, unde alela ALT este cu o unitate de repetare mai mică decât referința • too_few_supporting_reads – locație eliminată prin filtrare deoarece există prea puține citiri de susținere în proba tumorală • weak_evidence – scorul variantei somatice nu atinge pragul • systematic_noise – locație eliminată prin filtrare în baza dovezilor de zgomot sistematic în probele normale

Antet	Descriere
INFO (INFORMAȚII)	<p>Intrările posibile pentru INFO (Informații) includ:</p> <ul style="list-style-type: none"> • DP – profunzimea aproximativă a citirilor (informative și neinformative); este posibil ca unele citiri să fi fost eliminate prin filtrare în baza mapq etc. • END – poziție de final a intervalului. • FractionInformativeReads – fracția de citiri informative din totalul citirilor. • MQ – calitatea mapării RMS. • MQRankSum – scorul Z la testul sumei rangurilor Wilcoxon pentru calitățile de mapare Alt vs. Ref. • ReadPosRankSum – scorul Z la testul sumei rangurilor Wilcoxon pentru eroarea sistematică de poziție Alt vs. Ref. • AQ – scorul zgomotului sistemic. • hotspot – locație somatică cunoscută, utilizată pentru a crește încrederea la definire. • SOMATIC – cel puțin o variantă în această poziție este somatică.
FORMAT	<p>Coloana Format enumeră câmpuri separate de două puncte: De exemplu, GT:GQ. Printre câmpurile disponibile se numără:</p> <ul style="list-style-type: none"> • AD – profunzimi alelice (numărând doar citirile informative din totalul citirilor) pentru alelele de REF și ALT în ordinea enumerată. • AF – fracții de alele pentru alelele ALT în ordinea enumerată. • DP – profunzimea aproximativă a citirilor (sunt filtrate citirile cu MQ=255 sau cu perechi inadecvate). • F1R2 – numărul de citiri în orientarea perechii F1R2 care susține fiecare alelă. • F2R1 – numărul de citiri în orientarea perechii F2R1 care susține fiecare alelă. • GP – probabilități posterioare pe scara Phred pentru genotipuri, astfel cum este definit în specificația VCF. • GQ – calitatea genotipului. • GT – 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. • MB – statisticile componentelor pentru fiecare probă pentru a detecta decalajul dintre corespondenți. • PL – probabilitățile normalizate pe scara Phred pentru genotipuri, astfel cum este definit în specificația VCF. • PRI – probabilități posterioare pe scara Phred pentru genotipuri. • PS – 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. • SB – statisticile componentelor pentru fiecare probă care cuprind testul exact al lui Fisher pentru a detecta decalajul catenelor. • SQ – calitatea somatică.
SAMPLE (PROBĂ)	Coloana Sample (Probă) prezintă valorile specificate în coloana FORMAT.

Fișierele VCF genom

Fișierele VCF genom (*.gvcf.gz) respectă un set de convenții pentru reprezentarea tuturor siturilor din genom într-un format rezonabil de compact. Fișierele gVCF 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) de ./. indică absența unei definiții (no-call).

Retrimiteri analiză în coadă

Puteți retrimite analiza în coadă dacă analiza a fost oprită, dacă analiza nu a avut succes sau dacă doriți să reanalizați o rulare cu setări diferite. Pentru a retrimite analiza în coadă, parcurgeți următorii pași:

1. Din ecranul Run (Rulare), selectați fila Completed (Finalizate) și apoi selectați denumirea rulării ce va fi reanalizată.
Dacă a fost efectuată anterior retrimiteri analizei în coadă, selectați denumirea rulării pentru Parent Run (Rularea principală).
2. Din ecranul Run Details (Detaliile rulării), după Sequencing Information (Informații secvențiere), selectați **Requeue Analysis** (Retrimiteri analiză în coadă).
3. Selectați o opțiune:
 - Retrimiteri analizei în coadă fără modificări
 - Editarea setărilor rulării și retrimiteri analizei în coadă
 - Retrimiteri analizei în coadă cu o aplicație diferită
4. Confirmați că locația unde se află în prezent datele de secvențiere este furnizată în câmpul **Sequencing data file path** (Calea fișierului de date de secvențiere).

NOTĂ Calea către datele de secvențiere trebuie să se potrivească cu calea din setarea External Storage for Analysis (Spațiu extern de stocare pentru analiză). Consultați Ghid software pentru Illumina Run Manager for NextSeq 550Dx (document nr. 200025239) pentru informații despre schimbarea căii pentru spațiul extern de stocare.

5. Introduceți un motiv pentru reanalizare.
6. Selectați **Requeue Analysis** (Retrimiteri analiză în coadă).
7. Editați modificările dorite în Run Settings (Setări rulare), Sample Data (Date probă) și Analysis Settings (Setări analiză).
8. Selectați **Save** (Salvare). Analiza începe folosind parametrii de analiză curenți.

Asistență tehnică

Pentru asistență tehnică, contactați departamentul Asistență tehnică al Illumina.

Site web: www.illumina.com

E-mail: techsupport@illumina.com

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

Documentația produselor – disponibilă pentru descărcare de pe support.illumina.com.



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



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.

© 2023 Illumina, Inc. Toate drepturile rezervate.

illumina[®]