

DRAGEN for Illumina DNA Prep with Enrichment Dx

Documentația produsului NovaSeq 6000Dx

PROPRIETATEA ILLUMINA

Document nr. 200014776 v02

Septembrie 2022

A SE UTILIZA LA DIAGNOSTICAREA IN VITRO

Istoricul versiunilor

| Document | Data | Descrierea modificării |
|---------------|-----------------|---|
| 200014776 v02 | Septembrie 2022 | <p>A fost corectat formatul pentru fișierul manifest din text (*.txt) în BED (*.bed) la instrucțiunile pentru crearea rulării.</p> <p>Au fost corectate fișierele VCF cu consens în fișiere VCF în secțiunea cu fișierele de ieșire pentru analiză.</p> |
| 200014776 v01 | August 2022 | <p>S-a adăugat: Secțiunea Setări. Secțiunea Filtrarea zgomotului sistematic.</p> <p>Au fost actualizate instrucțiunile de creare a rulării pentru a include mai multe detalii.</p> <p>Au fost corectate greșelile de tipar și erorile gramaticale.</p> <p>S-a specificat faptul că instrucțiunile sunt destinate aplicației atunci când sunt utilizate cu Instrumentul NovaSeq 6000Dx.</p> <p>Au fost actualizate informațiile privind conținutul fișierului de ieșire VCF.</p> |
| 200014776 v00 | Martie 2022 | Versiunea inițială. |

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, 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).

© 2022 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.

Cuprins

| | |
|--|-----------|
| Istoricul versiunilor | ii |
| Prezentare generală | 1 |
| Metode de analiză | 1 |
| Crearea unei rulări | 4 |
| Setări | 5 |
| Rezultatele analizei | 7 |
| Fișiere FASTQ | 8 |
| Fișiere BAM | 8 |
| Fișiere VCF | 9 |
| Vizualizarea rezultatelor analizei | 15 |
| Asistență tehnică | 16 |

Prezentare generală

Aplicația DRAGEN™ for Illumina® DNA Prep with Enrichment Dx efectuează demultiplexarea, generarea FASTQ, maparea citirilor și alinierea la un genom de referință și definirea variantelor, în funcție de fluxul de lucru selectat pentru analiză.

Metode de analiză

DRAGEN for Illumina DNA Prep with Enrichment Dx efectuează demultiplexarea, generarea FASTQ, maparea citirii și alinierea la un genom de referință, în funcție de fluxul de lucru selectat:

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

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. 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

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 aliniere 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 (mişcare diagonală), o deleție (mişcare orizontală) sau o inserție (mişcare 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.

Valorile specifice alese pentru scoruri în acest algoritm indică modul de echilibrare, pentru o aliniere cu mai multe interpretări posibile, a posibilității de înregistrare a unui indel, spre deosebire de una sau mai multe SNP-uri sau preferința pentru o aliniere fără decupare. Valorile implicite ale scorului DRAGEN sunt rezonabile pentru alinierea citirilor de lungime moderată cu un genom de referință uman complet pentru aplicații cu definirea variantelor. Orice set de parametri de punctare Smith-Waterman reprezintă un model imprecis de erori de mutație genomică și de secvențiere. Valorile scorurilor de aliniere ajustate diferit pot fi mai potrivite pentru anumite aplicații.

Definirea variantelor DRAGEN Germline

Definitorul de variante mici DRAGEN Germline preia citirile de ADN mapate și aliniat ca date de intrare și numește SNP-urile și indeli printr-o combinație de detectare la nivel de coloană și asamblare *de novo* locală de haplotipuri.

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 în jurul coloanelor 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 aliniat 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.

În modul gVCF utilizat pentru definirea variantelor cu mai multe probe scalabile, Definitorul de variante mici DRAGEN Germline poate fi rulat pentru fiecare probă pentru a genera un fișier de definire a variantei genomice intermediare (gVCF). gVCF se poate utiliza apoi pentru genotiparea comună eficientă a mai multor probe, ceea ce permite procesarea incrementală rapidă a probelor și scalarea la dimensiuni mari de cohorte.

Deoarece Definitorul de variante mici DRAGEN Germline are algoritmi care îi permit să distingă în mod eficient erorile corelate de variantele reale, regulile de filtrare sunt foarte simple.

Definirea variantelor DRAGEN Somatic

Definitorul de variante mici DRAGEN Somatic preia citirile de ADN mapate și aliniat ca date de intrare și numește SNV-urile și indeli prin asamblarea *de novo* locală de haplotipuri într-o regiune activă.

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 în jurul coloanelor 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 aliniat 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 TLOD, Definitorul de variante mici DRAGEN Somatic scanează mai întâi în funcție de 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.

Crearea unei rulări

Utilizați pașii următori pentru a configura o rulare în Illumina Run Manager fie pe NovaSeq 6000Dx, fie utilizând un browser pe un computer din rețea. Datele probei pot fi introduse manual sau importând o fișă de probe.

Setări pentru aplicație și rulare

1. Din ecranul Runs (Rulări), selectați **Create Run** (Creare rulare).
2. Selectați aplicația DRAGEN for Illumina DNA Prep with Enrichment Dx, 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.
5. Asigurați-vă că Library Prep Kit (Setul de pregătire) a bibliotecii selectat este un set de pregătire a bibliotecii Illumina DNA Prep with Enrichment Dx.
6. Selectați setul de adaptoare de indexare dorit.
7. Introduceți Read Length (Lungimea de citire).
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.
8. **[Opțional]** Introduceți ID-ul unei eprubete din bibliotecă.
9. Selectați **Next** (Înainte).

Datele probei

Utilizați tabelul de pe ecranul Sample Data (Date probă) pentru a introduce manual informațiile despre probă. Alternativ, selectați **Import Samples** (Importare probe) pentru a încărca informațiile despre probe. Pentru informații privind importarea informațiilor despre probe, consultați [Importar probe la pagina 5](#).

Introducerea manuală a probelor

1. Introduceți un ID de probă unic în câmpul Sample ID (ID probă).
2. Utilizați **Plate - Well Position** (Placă - Poziție godeu) pentru a selecta poziția godeului.
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.
5. Selectați **Next** (Înainte).

Importar probe

Un șablon (*.csv) este disponibil pentru descărcare pe ecranul Sample Data (Date probă) atunci când se planifică o rulare în Illumina Run Manager folosind un browser pe un computer din rețea.

1. Selectați **Download Template** (Descărcare șablon) pentru a descărca un fișier CSV gol.
2. Din fișierul CSV, introduceți informațiile despre probă și salvați fișierul.
Fișierul CSV cu fișa de probă include următoarele coloane de date: Sample ID (ID probă), Plate - Well Position (Placă - Poziție godeu), **Opțional** Library Name (Denumire bibliotecă).
3. Selectați **Import Samples** (Importare probe) și navigați la locația fișierului CSV.
4. Selectați **Next** (Înainte).

Setări analiză

1. Selectați fluxul de lucru pentru analiză dorit:
 - Generare FASTQ
 - Generarea liniei germinale FASTQ și VCF pentru un flux de lucru Germline
 - Generarea somatică FASTQ și VCF pentru un flux de lucru somatic
2. **[Opțional]** Dacă doriți, selectați caseta de selectare **Generate ORA compressed FASTQs** (Generare FASTQ-uri comprimate ORA) pentru a activa compresia ORA pentru FASTQ.
3. **[Fluxuri de lucru pentru generarea VCF]** Utilizați meniul derulant **Manifest File Selection** (Selectare fișier manifest) pentru a selecta un fișier manifest.
Un fișier manifest reprezintă parametrul de intrare necesar pentru DRAGEN for Illumina DNA Prep with Enrichment Dx. Manifestul este un fișier BED delimitat cu tabulatori (*.bed), care specifică numele și locațiile regiunilor de referință vizate.
4. **[Fluxul de lucru pentru generarea FASTQ somatic și VCF]** Utilizați meniul derulant **Noise File Selection** (Selectare fișier zgomot) pentru a selecta un fișier de zgomot.
Un fișier BED cu nivel de zgomot specific unității poate fi specificat pentru filtrarea zgomotului sistematic. Pentru mai multe informații, consultați [Filtrarea zgomotului la pagina 6](#).
5. Selectați **Next** (Înainte).

Rulare revizuire

1. Pe ecranul Review (Revizuire), examinați informațiile introduse în ecranele 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

Selectați aplicația pe ecranul Applications (Aplicații) pentru a vizualiza setările actuale și 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.

Permisuni

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

Filtrarea zgomotului

Filtrarea sistematică a zgomotului este disponibilă atunci când se utilizează fluxul de lucru somatic.

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.

BED pentru zgomotul sistematic trebuie generat din probe normale. Se recomandă crearea unor fișiere de zgomot sistematic, care să fie specifice pregătirii bibliotecii, sistemului de secvențiere și panoului. Se recomandă utilizarea a aproximativ 50 de probe normale pentru generarea fișierelor de zgomot.

Rezultatele analizei

DRAGEN for Illumina DNA Prep with Enrichment Dx salvează următoarele informații în folderul de analiză. Numai fluxurile de lucru de linie germinală și somatice produc un PDF.

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

Fișierele de ieșire ale analizei

Următoarele fișiere de ieșire sunt generate de aplicație. Fișierele exacte generate depind de fluxul de lucru al analizei utilizat. Fișierele de ieșire se regăsesc în folderul analizei.

| Fișier de ieșire | Descriere |
|---|--|
| 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. |
| Fișiere BAM de aliniere (*.bam) | Conține citiri aliniate 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 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
- Un semn plus (+)
- Scorurile de calitate Phred într-un format codificat ASCII + 33

Identificatorul probei are următorul format.

```
@Instrument:IDRulare:IDCelulăFlux:Linie:Dală:X:Y
NumCitire:SemnalizareFiltru:0:NumărProbă
Exemplu:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAAA9#:<#<;<<<?????#=#
```

Fișiere BAM

Un fișier BAM (*.bam) este versiunea binară comprimată a unui fișier SAM (harta alinierii secvenței), care este folosită pentru a reprezenta secvențe aliniată 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 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 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. |
| 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 fișierelor VCF

| Antet | Descriere |
|--------------------|---|
| FILTER (FILTRU) | <p>Dacă se trece de toate filtrele, apare consemnat PASS (REUȘITĂ) în coloana pentru Filter (Filtru).</p> <p>Intrările posibile pentru FILTER (FILTRU) în fluxul de lucru Germline includ:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL (DRAGEN SNP QUAL mare) – se aplică dacă scorul QUAL pentru varianta SNP nu atinge pragul • DRAGENIndelHardQUAL (DRAGEN indel QUAL mare) – se aplică dacă scorul QUAL pentru varianta indel nu atinge pragul • LowDepth (profundzime redusă) – locație eliminată prin filtrare deoarece profundzimea de acoperire nu atinge pragul • LowGQ (GQ scăzut) – locație eliminată prin filtrare deoarece calitatea genotipului nu atinge pragul • PloidyConflict (conflict ploidie) – definirea genotipului de la definatorul de variante nu este în concordanță cu ploidia cromozomilor • base_quality (calitatea bazei) – locație eliminată prin filtrare deoarece calitatea mediană a bazei pentru citirile ALT în acest locus nu atinge pragul • filtered_reads (citiri filtrate) – locație eliminată prin filtrare deoarece o fracțiune prea mare de citiri a fost eliminată prin filtrare • fragment_length (lungime fragment) – 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 (profundzime redusă) – locație eliminată prin filtrare deoarece adâncimea de citire este prea scăzută • low_frac_info_reads (fracție de citiri informative scăzută) – locație eliminată prin filtrare deoarece fracțiunea de citiri informative este sub prag • low_normal_depth (profundzime redusă probă normală) – locație eliminată prin filtrare deoarece adâncimea de citire a probei normale este prea scăzută • long_indel (indel lung) – locație eliminată prin filtrare deoarece lungimea indelului este prea mare • mapping_quality (calitatea mapării) – locație eliminată prin filtrare deoarece calitatea mediană a mapării pentru citirile ALT în acest locus nu atinge pragul • multialelic (multialelic) – locație eliminată prin filtrare deoarece mai mult de două alele ALT trec de LOD tumorală • non_homref_normal (non_homref_normal) – locație eliminată prin filtrare deoarece genotipul probei normale nu este o referință homozigotă • no_reliable_supporting_read (lipsă citire de susținere fiabilă) – locație eliminată prin filtrare deoarece nu există o citire somatică de susținere fiabilă • panel_of_normals (panou de probe normale) – văzut în cel puțin o probă din panoul de vcf cu probe normale |

| Antet | Descriere |
|--------------------|--|
| FILTER (FILTRU) | <ul style="list-style-type: none">• read_position (poziție citire) – locație eliminată prin filtrare deoarece mediana distanțelor dintre începutul/sfârșitul citirii și acest locus este sub prag• RMxNRepeatRegion (regiune repetată RMxN) – locație eliminată prin filtrare deoarece întreaga sau o parte a alelei variantei este o repetare a referinței• strand_artifact (artefacte catenă) – locație eliminată prin filtrare din cauza unui decalaj sever al catenelor• str_contraction (constracție catene) – 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 (prea puține citiri de susținere) – locație eliminată prin filtrare deoarece există prea puține citiri de susținere în proba tumorală• weak_evidence (dovadă slabă) – scorul variantei somatice nu atinge pragul |

| Antet | Descriere |
|------------------------------------|---|
| FILTER (FILTRU) (continuare) | <p>Intrările posibile pentru FILTER (FILTRU) în fluxul de lucru somatic includ:</p> <ul style="list-style-type: none"> • base_quality (calitatea bazei) – locație eliminată prin filtrare deoarece calitatea mediană a bazei pentru citirile ALT în acest locus nu atinge pragul • filtered_reads (citiri filtrate) – locație eliminată prin filtrare deoarece o fracțiune prea mare de citiri a fost eliminată prin filtrare • fragment_length (lungime fragment) – 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 (profundime redusă) – locație eliminată prin filtrare deoarece adâncimea de citire este prea scăzută • low_frac_info_reads (fracție de citiri informative scăzută) – locație eliminată prin filtrare deoarece fracțiunea de citiri informative este sub prag • low_normal_depth (profundime redusă probă normală) – locație eliminată prin filtrare deoarece adâncimea de citire a probei normale este prea scăzută • long_indel (indel lung) – locație eliminată prin filtrare deoarece lungimea indelului este prea mare • mapping_quality (calitatea mapării) – locație eliminată prin filtrare deoarece calitatea mediană a mapării pentru citirile ALT în acest locus nu atinge pragul • multiallelic (multiallelic) – locație eliminată prin filtrare deoarece mai mult de două alele ALT trec de LOD tumorală • non_homref_normal (non_homref_normal) – locație eliminată prin filtrare deoarece genotipul probei normale nu este o referință homozigotă • no_reliable_supporting_read (lipsă citire de susținere fiabilă) – locație eliminată prin filtrare deoarece nu există o citire somatică de susținere fiabilă • panel_of_normals (panou de probe normale) – văzut în cel puțin o probă din panoul de vcf cu probe normale • read_position (poziție citire) – locație eliminată prin filtrare deoarece mediana distanțelor dintre începutul/sfârșitul citirii și acest locus este sub prag • RMxNRepeatRegion (regiune repetată RMxN) – locație eliminată prin filtrare deoarece întreaga sau o parte a alelei variantei este o repetare a referinței • strand_artifact (artefacte catenă) – locație eliminată prin filtrare din cauza unui decalaj sever al catenelor • str_contraction (contractie catene) – 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 (prea puține citiri de susținere) – locație eliminată prin filtrare deoarece există prea puține citiri de susținere în proba tumorală • weak_evidence (dovadă slabă) – scorul variantei somatice nu atinge pragul |

| Antet | Descriere |
|------------------------------------|--|
| FILTER (FILTRU) (continuare) | <ul style="list-style-type: none"> • systematic_noise (zgomot sistematic) – locație eliminată prin filtrare în baza dovezilor de zgomot sistematic în probele normale |
| INFO (INFORMAȚII) | <p>Intrările posibile pentru INFO (INFORMAȚII) în fluxul de lucru al liniei germinale 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 (decalaj R2_5P) – 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 (CAPĂT) – poziție de final a intervalului. • FractionInformativeReads (fracție citiri informative) – fracția de citiri informative din totalul citirilor. • MQ – calitatea mapării RMS. • MQRankSum (suma rangurilor MQ) – scorul Z la testul sumei rangurilor Wilcoxon pentru calitățile de mapare Alt vs. Ref. • ReadPosRankSum (suma rangurilor poziției de citire) – 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. <p>Intrările posibile pentru INFO (INFORMAȚII) în fluxul de lucru somatic 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 (CAPĂT) – poziție de final a intervalului. • FractionInformativeReads (fracție citiri informative) – fracția de citiri informative din totalul citirilor. • MQ – calitatea mapării RMS. • MQRankSum (suma rangurilor MQ) – scorul Z la testul sumei rangurilor Wilcoxon pentru calitățile de mapare Alt vs. Ref. • ReadPosRankSum (suma rangurilor poziției de citire) – scorul Z la testul sumei rangurilor Wilcoxon pentru eroarea sistematică de poziție Alt vs. Ref. • AQ – scorul zgomotului sistemic. |

| Antet | Descriere |
|----------------------|---|
| INFO (INFORMAȚII) | <ul style="list-style-type: none"> • 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.</p> <p>Printre câmpurile disponibile în fluxul de lucru Germline 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. |

| Antet | Descriere |
|----------------|---|
| FORMAT | <ul style="list-style-type: none"> • 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. <p>Printre câmpurile disponibile în fluxul de lucru somatic 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. |

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).

Vizualizarea rezultatelor analizei

Rulările în curs în prezent sunt afișate în fila Active. Rulările finalizate sunt afișate în fila Completed (Finalizate). Consultați [Documentația produsului NovaSeq 6000Dx \(document nr. 200010105\)](#) pentru mai multe informații despre vizualizarea rezultatelor.

Asistență tehnică

Pentru asistență tehnică, contactați departamentul Asistență tehnică al Illumina.

Site web: www.illumina.com
E-mail: techsupport@illumina.com

Numere de telefon pentru asistență tehnică Illumina

| Regiune | Număr de telefon gratuit | Internațional |
|------------------|--------------------------|-------------------|
| Australia | +61 1800 775 688 | |
| Austria | +43 800 006249 | +43 1 9286540 |
| Belgia | +32 800 77 160 | +32 3 400 29 73 |
| Canada | +1 800 809 4566 | |
| China | | +86 400 066 5835 |
| Danemarca | +45 80 82 01 83 | +45 89 87 11 56 |
| Finlanda | +358 800 918 363 | +358 9 7479 0110 |
| Franța | +33 8 05 10 21 93 | +33 1 70 77 04 46 |
| Germania | +49 800 101 4940 | +49 89 3803 5677 |
| Hong Kong, China | +852 800 960 230 | |
| India | +91 8006500375 | |
| Indonezia | | 0078036510048 |
| Irlanda | +353 1800 936608 | +353 1 695 0506 |
| Italia | +39 800 985513 | +39 236003759 |
| Japonia | +81 0800 111 5011 | |
| Malaezia | +60 1800 80 6789 | |
| Țările de Jos | +31 800 022 2493 | +31 20 713 2960 |
| Noua Zeelandă | +64 800 451 650 | |
| Norvegia | +47 800 16 836 | +47 21 93 96 93 |
| Filipine | +63 180016510798 | |
| Singapore | 1 800 5792 745 | |
| Coreea de Sud | +82 80 234 5300 | |
| Spania | +34 800 300 143 | +34 911 899 417 |

| Regiune | Număr de telefon gratuit | Internațional |
|----------------------------|--------------------------|------------------|
| Suedia | +46 2 00883979 | +46 8 50619671 |
| Elveția | +41 800 200 442 | +41 56 580 00 00 |
| Taiwan, China | +886 8 06651752 | |
| Thailanda | +66 1800 011 304 | |
| Regatul Unit | +44 800 012 6019 | +44 20 7305 7197 |
| Statele Unite ale Americii | +1 800 809 4566 | +1 858 202 4566 |
| Vietnam | +84 1206 5263 | |

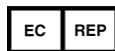
Fișe cu date de securitate (SDS) – disponibile pe site-ul web Illumina la adresa support.illumina.com/sds.html.

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

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