

# DRAGEN za Illumina DNK Prep with Enrichment Dx

Dokumentacija proizvoda za NovaSeq 6000Dx

## Povijest revizija

Dokument	Datum	Opis promjene
200014776 v02	Rujan 2022.	Ispravljeni format za datoteku manifesta od teksta (*.txt) do BED-a (*.bed) u uputama za izradu ciklusa obrade. Ispravljene konsenzusne VCF datoteke u VCF datoteke u odjeljku izlaznih rezultata analize.
200014776 v01	Kolovoz 2022.	Dodano: Odjeljak Postavke. Odjeljak Filtriranje sustavne buke. Ažurirane su upute za izradu ciklusa obrade kako bi se uključilo više pojedinosti. Ispravljene su tipografske i gramatičke pogreške. Navedeno je da su upute namijenjene aplikaciji kada se upotrebljava s instrumentom NovaSeq 6000Dx. Ažurirane su informacije o sadržaju izlazne VCF datoteke.
200014776 v00	Ožujak 2022.	Početno izdanje.

Ovaj dokument i njegov sadržaj vlasništvo su tvrtke Illumina, Inc. i njezinih povezanih društava („Illumina“) te su namijenjeni isključivo za ugovornu upotrebu klijentima u vezi s proizvodom(ima) opisanu u njemu(ima). Dokument i njegov sadržaj ne smiju se upotrebljavati ni distribuirati ni u koju drugu svrhu niti se smiju na neki drugi način prenositi, otkrivati ili reproducirati bez prethodnog pisanog odobrenja tvrtke Illumina. Illumina ovim dokumentom ne prenosi nikakve licence zaštićene svojim pravom na patent, žig, autorskim pravom ili običajnim pravom ni slična prava bilo koje treće strane.

Kvalificirano i odgovarajuće obučeno osoblje mora se strogo i bez iznimki pridržavati uputa u ovom dokumentu da bi se zajamčila pravilna i sigurna upotreba proizvoda opisana u njemu. Prije upotrebe proizvoda nužno je s razumijevanjem pročitati cjelokupan sadržaj dokumenta.

AKO UPUTE U DOKUMENTU NE PROČITATE U CIJELOSTI TE IH SE NE PRIDRŽAVATE BEZ IZNIMKI, MOŽE DOĆI DO OŠTEĆENJA PROIZVODA, OZLJEDA KORISNIKA ILI DRUGIH OSOBA I DO OŠTEĆENJA DRUGE IMOVINE TE SE TIME PONIŠTAVAJU SVA JAMSTVA ZA PROIZVODE.

ILLUMINA NE PREUZIMA ODGOVORNOST ZA ŠTETE NASTALE USLIJED NEPRAVILNE UPOTREBE PROIZVODA KOJI SU OPISANI U OVOM DOKUMENTU (UKLJUČUJUĆI DIJELOVE TIH PROIZVODA I SOFTVER).

© 2022. Illumina, Inc. Sva prava pridržana.

Svi su žigovi vlasništvo tvrtke Illumina, Inc. ili svojih vlasnika. Konkretno informacije o žigovima potražite na adresi [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).

# Sadržaj

Povijest revizija .....	ii
Pregled .....	1
Metode analize .....	1
Izrada obrade .....	4
Postavke .....	5
Rezultati analize .....	7
FASTQ datoteke .....	8
BAM datoteke .....	8
VCF datoteke .....	9
Prikaz rezultata analize .....	14
<b>Tehnička pomoć .....</b>	<b>15</b>

# Pregled

Aplikacija DRAGEN™ za Illumina® DNA Prep with Enrichment Dx provodi demultipleksiranje, generiranje FASTQ-a, očitavanje mapiranja i usklađivanje s referentnim genomom i očitavanje varijanti, ovisno o odabranom tijeku analize.

## Metode analize

DRAGEN za Illumina DNK Prep with Enrichment Dx provodi demultipleksiranje, generiranje FASTQ-a, mapiranje očitavanja i usklađivanje s referentnim genomom ovisno o odabranom tijeku rada:

- FASTQ generacija
- Germline FASTQ i VCF generiranje
- Somatic FASTQ i VCF generiranje

## FASTQ generiranje

Sastavljeni slijedovi spremaju se u FASTQ datoteke po uzorku. FASTQ datoteke tekstualne su datoteke koje sadrže podatke o sekvenciranju i rezultate kvalitete samo za jedan uzorak. Za svaki uzorak generiraju se zasebne FASTQ datoteke po stazi protočne članke, po očitavanju sekvenciranja. Naziv uzorka naveden tijekom postavljanja obrade uključen je u naziv FASTQ datoteke. Datoteke FASTQ primarne su ulazne datoteke za usklađivanje. Prvi korak FASTQ generiranja je demultipleksiranje. Demultipleksiranjem se dodjeljuju klasteri koji prolaze filter uzorku usporedbom svake sekvence očitavanja indeksa sa sekvencama indeksa navedenim za obradu. U tom se koraku ne gledaju vrijednosti kvalitete. Očitavanja indeksa prepoznaju se u sljedećim koracima:

- Uzorci se numeriraju počevši od 1 na temelju redoslijeda kojim su navedeni za obradu.
- Broj uzorka 0 rezerviran je za klaster koji nisu dodijeljeni uzorku.
- Klasteri se dodjeljuju uzorku kad se sekvenca indeksa točno podudara ili kad se nađe maksimalno jedno nepodudaranje po očitavanju indeksa.

Softver uključuje kompresiju ORA za komprimiranje FASTQ datoteka. Kada se koristi format ORA (\*.ora), md5 checksum datoteka FASTQ sadržaja čuva se nakon ciklusa kompresije i dekompresije kako bi se osigurala kompresija bez gubitka podataka.

## Mapiranje DNK i poravnanje

Prva faza mapiranja je generiranje sjemena iz očitavanja, a zatim traženje točnih podudaranja u referentnom genomu. Ti se rezultati zatim rafiniraju pokretanjem cijelog Smith-Watermanova algoritama poravnanja na mjestima s najvećom gustoćom podudaranja sjemena. Ovaj dobro dokumentirani algoritam radi usporedbom svakog položaja očitavanja sa svim položajima kandidata u referentnom genomu. Te usporedbe odgovaraju matrici potencijalnih poravnanja između očitavanja i referentnog

genoma. Za svaki od ovih položaja poravnanja kandidata, Smith-Watermanov algoritam generira rezultate koji se koriste kako bi se procijenilo je li najbolje poravnanje koje prolazi kroz tu ćeliju matrice nju doseglo podudaranjem ili neslaganjem nukleotida (dijagonalno pomicanje), brisanjem (vodoravno pomicanje) ili umetanjem (okomito kretanje). Podudaranje između očitavanja i reference daje bonus na rezultat, a nepodudaranje ili indel nameće kaznu. Ukupni najviši put ocjenjivanja kroz matricu je odabrano poravnanje.

Specifične vrijednosti odabrane za ocjene u ovom algoritmu ukazuju na to kako uravnotežiti, poravnanje s više mogućih tumačenja, mogućnost indela za razliku od jednog ili više SNP-ova ili sklonost poravnanju bez odrezivanja. Zadane vrijednosti DRAGEN ocjenjivanja razumne su za poravnanje očitavanja umjerene duljine s cijelim ljudskim referentnim genomom za primjene očitavanja varijanti. Bilo koji skup Smith-Watermanovih parametara za ocjenjivanje predstavlja neprecizni model genomske mutacije i pogrešaka sekvenciranja. Različito prilagođene vrijednosti ocjenjivanja poravnanja mogu biti prikladnije za neke primjene.

## Očitavanje DRAGEN Germline Variant

DRAGEN Germline Small Variant Caller mapira i poravnava DNK očitavanja kao ulazna i prepoznaje SNP-ove i indele kombinacijom očitavanja po stupcima i lokalnog *novog* skupa haplotipova.

Referentna područja koja se mogu očitati najprije se identificiraju uz dovoljnu pokrivenost poravnanja. Unutar tih referentnih regija, brzo skeniranje sortiranih očitavanja prepoznaje aktivna područja koja su centrirana oko stupaca nakupljanja s dokazima o varijanti. Aktivne regije predstavljene su s dovoljno konteksta za pokrivanje značajnog sadržaja nerefereentnog sadržaja u blizini. Ako postoje dokazi o indelima, aktivne regije dobivaju dodatnu osnovu.

Usklađena očitavanja odrezuju se unutar svake aktivne regije i sastavljaju u De Bruijnov grafikon. Rubovi odrezanih očitavanja ponderirani su brojevima promatranja, s referentnom sekvencom kao okosnicom. Nakon čišćenja grafikona i pojednostavljenja, svi putovi pada ili porasta na grafikonu se izdvajaju kao haplotipovi kandidata. Svaki haplotip je Smith-Waterman poravnat s referentnim genomom kako bi se identificirale varijante koje predstavlja. Ovaj skup događaja može se pojačati otkrivanjem na temelju položaja. Za svaki par očitanih haplotipa, vjerojatnost  $P(r|H)$  promatranja očitavanja, pod pretpostavkom da je haplotip pravi početni uzorak, procjenjuje se pomoću skrivenih Markovljevih modela (HMM).

Skeniranje po referentnom položaju u odnosu na aktivno područje, genotipovi kandidata formiraju se iz diploidnih kombinacija varijantnih događaja (SNP-ovi ili indeli). Za svaki događaj (uključujući referentni), uvjetna vjerojatnost  $P(r|e)$  promatranja svakog preklapanja očitavanja procjenjuje se kao maksimum  $P(r|H)$  za haplotipove koji podupiru događaj. Oni se kombiniraju u uvjetnu vjerojatnost  $P(r|e1e2)$  za genotip (par događaja) i množe kako bi se dobila uvjetna vjerojatnost  $P(R|e1e2)$  promatranja cijelog očitanih nakupljanja. Pomoću Bayesove formule izračunava se posteriorna vjerojatnost  $P(e1e2|R)$  svakog diploidnog genotipa i očitava se pobjednik.

U načinu rada gVCF koji se koristi za skalabilno očitavanje varijanti s većim brojem uzoraka, DRAGEN Germline Small Variant Caller može se pokrenuti obradu po uzorku za generiranje sporedne datoteke

očitavanja genomske varijante (gVCF). gVCF se zatim može koristiti za učinkovitu udruženu genotipizaciju više uzoraka, što omogućuje brzu postupnu obradu uzoraka i skaliranje do velikih veličina kohorte.

Budući da DRAGEN Germline Small Variant Caller ima algoritme koji omogućuju učinkovito razlikovanje povezanih pogrešaka od pravih varijanti, pravila filtriranja su vrlo jednostavna.

## Očitavanje DRAGEN Somatic Variant

DRAGEN Somatic Small Variant Caller mapira i poravnava DNK očitavanja kao ulazna i prepoznaje SNV-ove i indele putem lokalnog *novog* skupa haplotipova u aktivnom području.

Referentna područja koja se mogu očitati najprije se identificiraju uz dovoljnu pokrivenost poravnanja. Unutar tih referentnih regija, skeniranje sortiranih očitavanja prepoznaje aktivne regije koje su centrirane oko stupaca nakupljanja s dokazima o varijanti u očitavanjima tumora. Aktivne regije predstavljene su s dovoljno konteksta za pokrivanje značajnog sadržaja nereferentnog sadržaja u blizini. Ako postoje dokazi o indelima, aktivne regije dobivaju dodatnu osnovu.

Usklađena očitavanja odrezuju se unutar svake aktivne regije i sastavljaju u De Bruijnov grafikon. Rubovi odrezanih očitavanja ponderirani su brojevima promatranja, s referentnom sekvencom kao okosnicom. Nakon čišćenja grafikona i pojednostavljenja, svi putevi pada ili porasta na grafikonu se izdvajaju kao haplotipovi kandidata. Svaki haplotip je Smith-Waterman poravnat s referentnim genomom kako bi se identificirale varijante koje predstavlja. Za svaki par očitanih haplotipa, vjerojatnost  $P(r|H)$  promatranja očitavanja, procjenjuje se pomoću skrivenih Markovljevih modela (HMM) pod pretpostavkom da je haplotip pravi početni uzorak.

Da bi se odredio TLOD rezultat, DRAGEN Somatic Small Variant Caller prvo skenira u skladu s referentnim položajem za svaki somatski događaj kandidata, kao i referentnim događajem u aktivnoj regiji. Uvjetna vjerojatnost  $P(r|e)$  promatranja svakog očitavanja preklapanja procjenjuje se kao maksimum  $P(r|H)$  za haplotipove koji podupiru događaj. Oni se kombiniraju u uvjetnu vjerojatnost  $P(r|E)$  za hipotezu događaja, E, uključujući mješavinu referentnog i somatskog alela kandidata u rasponu mogućih frekvencija alela i množe se kako bi se dobila uvjetna vjerojatnost  $P(R|E)$  promatranja cijelog očitavanja nakupljanja. Od tamo se izračunava TLOD rezultat kao dokaz da je alel ALT prisutan u uzorku tumora na određenom lokusu.

# Izrada obrade

Pomoću sljedećih koraka postavite obradu u Illumina Run Manager na NovaSeq 6000Dx ili pomoću preglednika na umreženom računalu. Podaci o uzorku mogu se unijeti ručno ili uvozom lista s uzorcima.

## Postavke aplikacije i obrade

1. Na zaslonu Runs (Obrade) odaberite **Create Run** (Izradi obradu).
2. Odaberite aplikaciju DRAGEN za Illumina DNK Prep with Enrichment Dx, a zatim odaberite **Next** (Dalje).
3. Na zaslonu Run Settings (Postavke obrade) unesite naziv obrade. Naziv obrade prepoznaje obradu od sekvenciranja do analize.
4. **[Neobavezno]** Unesite opis obrade da biste je dodatno prepoznali.
5. Provjerite je li odabrani komplet za pripremu knjižnice Illumina DNA Prep with Enrichment Dx komplet za pripremu knjižnice.
6. Odaberite željeni komplet prilagodnika indeksa.
7. Unesite duljinu očitavanja.  
Očitavanje 1 i očitavanje 2 imaju zadanu vrijednost od 151 ciklusa.  
Indeks 1 i indeks 2 imaju fiksnu vrijednost od 10 ciklusa.
8. **[Neobavezno]** Unesite ID epruvete knjižnice.
9. Odaberite **Next** (Dalje).

## Podaci o uzorku

Pomoću tablice na zaslonu Sample Data (Podaci o uzorku) ručno unesite podatke o uzorku. Alternativno, odaberite **Import Samples** (Uvoz uzoraka) za prijenos informacija o uzorku. Informacije o uvozu informacija o uzorku potražite u odjeljku [Uvoz uzoraka na stranici 5](#).

### Ručni unos uzoraka

1. U polje Sample ID (ID uzorka) unesite jedinstveni ID uzorka.
2. Pomoću **Plate - Well Position** (Pločica - položaj jažice) odaberite položaj jažice.  
Polja i7 indeksa, indeksa 1, i5 indeksa i indeksa 2 automatski se popunjavaju.
3. **[Neobavezno]** Unesite naziv knjižnice.
4. Dodajte retke i po potrebi ponovite korake **1 – 3** dok svi uzorci ne budu dodani u tablicu.
5. Odaberite **Next** (Dalje).



## Uvoz uzoraka

Predložak (\*.csv) dostupan je za preuzimanje na zaslonu Sample Data (Podaci o uzorcima) prilikom planiranja obrade u Illumina Run Manager pomoću preglednika na umreženom računalu.

1. Odaberite **Download Template** (Preuzmi predložak) za preuzimanje prazne CSV datoteke.
2. U CSV datoteku unesite informacije o uzorku i spremite datoteku.  
CVS datoteka s listom s uzorcima sadrži sljedeće stupce s podacima: Sample ID (ID uzorka), Plate - Well Position (Pločica - položaj jažice), **Neobavezno** Library Name (Naziv knjižnice).
3. Odaberite **Import Samples** (Uvezi uzorke) i potražite lokaciju CSV datoteke.
4. Odaberite **Next** (Dalje).

## Postavke analize

1. Odaberite željeni tijek rada analize:
  - FASTQ generacija
  - Germline FASTQ i VCF generiranje za za tijek rada sa spolnim stanicama
  - Somatic FASTQ i VCF generiranje za somatski tijek rada
2. **[Neobavezno]** Ako želite, odaberite potvrdni okvir **Generate ORA compressed FASTQs** (Generiraj komprimirani FASTQ na formate ORA) kako biste omogućili kompresiju FASTQ u formatu ORA.
3. **[VCF generiranje tijekom rada]** Upotrijebite padajući izbornik **Manifest File Selection** (Odabir datoteke manifesta) za odabir datoteke manifesta.  
Datoteka manifesta je potreban unos za DRAGEN za Illumina DNK Prep with Enrichment Dx. Manifest je BED (\*.bed) datoteka s podacima odvojenim tabulatorom koja navodi nazive i lokacije ciljanih referentnih regija.
4. **[Tijek rada za Somatic FASTQ i VCF generiranje]** Na padajućem izborniku **Noise File Selection** (Odabir datoteke buke) odaberite datoteku s bukom.  
Može se odrediti BED datoteka s razinom buke specifičnom za lokaciju za filtriranje sustavne buke. Dodatne informacije potražite u odjeljku [Filtriranje buke na stranici 6](#).
5. Odaberite **Next** (Dalje).

## Obrada Pregled

1. Na zaslonu Review (Pregled) pregledajte informacije unesene na zaslonima Run Settings (Postavke obrade), Sample Data (Podaci o uzorcima) i Analysis Settings (Postavke analize).
2. Odaberite **Save** (Spremi).  
Obrada se sprema na kartici Planned (Planirano) na zaslonu Runs (Obrade).

## Postavke

Na zaslonu Applications (Aplikacije) odaberite aplikaciju za prikaz trenutnih postavki i promjenu postavki.

## Konfiguracija

Zaslon konfiguracije prikazuje sljedeće postavke aplikacije:

- **Library Prep Kits** (Kompleti za pripremu knjižnice)— Prikazuje zadani komplet za pripremu knjižnice za aplikaciju. Ova se postavka ne može promijeniti.
- **Index Adapter Kits** (Kompleti prilagodnika indeksiranja)— Prikazuje zadani komplet prilagodnika indeksiranja za aplikaciju. Ova se postavka ne može promijeniti.
- **Read Lengths** (Dužine očitavanja)— Duljine očitavanja za aplikaciju postavljene su prema zadanim postavkama na 151, ali se mogu promijeniti tijekom izrade ciklusa obrade.
- **Manifest and Noise Files** (Datoteke manifesta i buke)— Prenesite i promijenite postavke za datoteke manifesta i buke.
  - Odaberite **Upload File** (Prijenos datoteke) za prijenos datoteka za korištenje u analizi.
  - Odaberite **Default** (Zadano) izborni gumb za postavljanje datoteke kao zadane datoteke manifesta ili buke odabrane tijekom izrade ciklusa obrade kada je odabrana aplikacija.
  - Odaberite potvrdni okvir **Enabled** (Omogućeno) da biste na padajućem izborniku postavili datoteku za prikaz tijekom izrade ciklusa obrade.

## Dozvole

Upotrijebite potvrdne okvire na zaslonu Permissions (Dozvole) da biste upravljali pristupom korisnika aplikaciji.

## Filtriranje buke

Filtriranje sustavne buke dostupno je prilikom korištenja somatskog tijeka rada. Filtar se može koristiti u Tumor-Normal (Tumorsko-normalnom) načinu rada, ali je posebno koristan samo za obrade tumora gdje podudarna normalna vrijednost nije dostupna.

Sustavna buka BED-a treba se generirati iz normalnih uzoraka. Preporučuje se izrada sustavnih datoteka buke koje su specifične za pripremu knjižnice, sustav za sekvenciranje i ploču. Preporučuje se koristiti oko 50 normalnih uzoraka za generiranje datoteke s bukom.

## Rezultati analize

DRAGEN za Illumina DNK Prep with Enrichment Dx sprema sljedeće informacije u mapu analize. Samo spolni i somatski tijekovi rada generiraju izvješća u obliku PDF-a.

- Korištena datoteka manifesta
- Verzija softvera
- ID-ovi uzoraka
- Ukupne usklađene baze
- Postotak usklađenih očitavanja po uzorku
- Broj SNV-ova očitanih po uzorku
- Broj indela očitanih po uzorku
- Statistika pokrivenosti

### Izlazne datoteke analize

Aplikacija generira sljedeće izlazne datoteke. Koje će se točno datoteke generirati ovise o tome koji se tijek rada analize koristi. Izlazne datoteke nalaze se u mapi analize.

Izlazna datoteka	Opis
FASTQ (*.fastq.gz ili *.fastq.ora)	Sporedne datoteke koje sadrže očitavanja baza za koja postoji ocjena kvalitete. FASTQ datoteke primarne su ulazne datoteke u koraku usklađivanja. Ako je odabrana kompresija ORA, naziv datoteke će to odražavati.
Usklađivanje BAM datoteka (*.bam)	Sadrže usklađena očitavanja za određeni uzorak.
Genomske VCF datoteke (*.gvcf.gz)	Sadrži genotip za svaki položaj, očitani kao varijanta ili kao referenca.
VCF datoteke (*.vcf.gz)	Sadrže varijante očitane na svakom položaju.
Izvješće s mjernim podacima obrade (*.csv)	Sadrži mjerne podatke kvalitete obrade, uključujući ukupnu iskoristivost i ocjenu kvalitete za Q30.

## FASTQ datoteke

FASTQ (\*.fastq.gz, \*.fastq.ora) je format datoteke utemeljen na tekstu koji sadrži očitavanja baza i vrijednosti kvalitete po očitavanju. Svaka datoteka sadrži sljedeće informacije:

- Identifikator uzorka
- sekvence
- znaka plus (+)
- Phred bodovanja kvalitete u formatu kodiranom prema standardu ASCII + 33

Identifikator uzorka formatiran je na sljedeći način.

```
@Instrument:IDObrade:IDProtočneJedinice:Staza:Pločica:X:Y
BrojČitanja:ZastavicaFiltrira:0:BrojUzorka
Primjer:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAAA9#:<#<;<<<????#=#
```

## BAM datoteke

BAM datoteka (\*.bam) je komprimirana binarna verzija SAM (mapa poravnanja sekvenci) datoteke koja se upotrebljava za predstavljanje usklađenih sekvenci veličine do 128 Mb. BAM datoteke upotrebljavaju oblik imenovanja datoteka `SampleName_S#.bam`, pri čemu je # broj uzorka određen redoslijedom kojim su uzorci navedeni za obradu. U multinodnom načinu rada, S# je postavljen na S1, bez obzira na redoslijed uzorka.

BAM datoteke sadrže odjeljak zaglavlja i odjeljak usklađivanja:

- Zaglavlje – sadrži informacije o cijeloj datoteci kao što su naziv uzorka, dužina uzorka i metoda usklađivanja. Usklađivanja u odjeljku usklađivanja povezana su s određenim informacijama u odjeljku zaglavlja.
- Usklađivanja – sadrži naziv očitavanja, sekvencu očitavanja, kvalitetu očitavanja, informacije o usklađivanju i prilagođene oznake. Naziv očitavanja obuhvaća kromosom, početnu koordinatu, kvalitetu usklađivanja i podudarni niz opisnika.

Sekcija usklađivanja obuhvaća sljedeće informacije za svako očitavanje ili par očitavanja:

- AS: kvaliteta usklađivanja s uparenim krajevima.
- RG: grupa očitavanja koja upućuje na broj očitavanja za određeni uzorak.
- BC: oznaka crtičnog koda koja upućuje na ID demultipleksiranog uzorka povezanog s očitanjem.
- SM: kvaliteta usklađivanja s jednim krajem.

- XC: podudarni niz opisnika.
- XN: oznaka naziva amplikona koja bilježi ID amplikona povezanog s očitanim datotekama indeksa BAM (\*.bam.bai) daje indeks odgovarajuće BAM datoteke.

## VCF datoteke

Format očitavanja varijante (\*.vcf) sadrži informacije o varijantama pronađenim na određenim mjestima u referentnom genomu.

Zaglavlje VCF datoteke sastoji se od verzije formata datoteke VCF i verzije očitavanja varijante te navodi oznake korištene u ostatku datoteke. VCF zaglavlje obuhvaća i datoteku referentnog genoma te BAM datoteku. Zadnji redak zaglavlja sadrži naslove stupaca za retke podataka. Svaki redak podataka u VCF datoteci sadrži informacije o jednoj varijanti.

Tablica 1 Zaglavlja VCF datoteka

Zaglavlje	Opis
CHROM	Kromosom referentnog genoma. Kromosomi se pojavljuju istim redoslijedom kao u referentnoj datoteci FASTA.
POS	Jednobazni položaj varijante u referentnom kromosomu. Za jednonukleotidne varijante (SNV-ove), ovaj je položaj referentna baza s varijantom. Za indele, ovaj je položaj referentna baza koja je neposredno ispred varijante.
ID	Rs (referentni SNP) broj varijante dobiven iz datoteke <code>dbSNP.txt</code> , ako je to primjenljivo. Ako na tom mjestu postoji više rs brojeva, popis je odijeljen točkama-zarezima. Ako na tom mjestu ne postoji dbSNP unos, primjenjuje se oznaka ('.') za vrijednost koja nedostaje.
REF	Referentni genotip. Primjerice, brisanje jednog T predstavljeno je kao referentni TT i alternativni T. Jedna varijanta nukleotida A na T predstavljena je kao referentni A i alternativni T.
ALT	Aleli koji se razlikuju od referentnog očitavanja. Primjerice, umetanje jednog T predstavljeno je kao referentni A i alternativni AT. Jedna varijanta nukleotida A na T predstavljena je kao referentni A i alternativni T.
QUAL	Phred bodovanje kvalitete koje dodjeljuje očitavatelj varijante. Veći broj bodova upućuje na veću pouzdanost varijante i manju mogućnost pogrešaka. Za bodovanje kvalitete Q procijenjena vjerojatnost pogreške iznosi $10^{-Q/10}$ . Primjerice, skup od Q30 očitavanja ima omjer pogrešaka od 0,1 %. Mnogi očitavatelji varijanti boduju kvalitetu na temelju svojih statističkih modela, koji visoko koreliraju s uočenim omjerom pogrešaka.

Tablica 2 Oznake u VCF datoteci

Zaglavlje	Opis
FILTAR	<p>Ako je datoteka prošla sve filtre, u stupcu filtra piše PASS (PROLAZ). Mogući FILTAR unosi za tijek rada sa spolnim stanicama uključuju:</p> <ul style="list-style-type: none"> <li>• <b>DRAGENSnpHardQUAL</b>—Primjenjuje se ako varijanta SNP QUAL rezultata ne zadovoljava granične vrijednosti</li> <li>• <b>DRAGENIndelHardQUAL</b>—Primjenjuje se ako indel varijanta QUAL rezultata ne zadovoljava granične vrijednosti</li> <li>• <b>LowDepth</b>—Lokacija je filtrirana jer dubina pokrivenosti ne zadovoljava granične vrijednosti</li> <li>• <b>LowGQ</b>—Centar filtriran jer kvaliteta genotipa ne zadovoljava prag</li> <li>• <b>PloidyConflict</b>—Očitavanje genotipa od pozivatelja varijanti nije u skladu s ploidiom kromosoma</li> <li>• <b>base_quality</b>—Lokacija je filtrirana jer srednja osnovna kvaliteta alt očitavanja na tom lokusu ne zadovoljava granične vrijednosti</li> <li>• <b>filtered_reads</b>—Lokacija je filtrirana jer je prevelik dio očitavanja filtriran</li> <li>• <b>fragment_length</b>—Lokacija je filtrirana jer apsolutna razlika između srednje duljine fragmenta alt očitavanja i srednje duljine fragmenta ref očitavanja na tom lokusu prelazi granične vrijednosti</li> <li>• <b>low_deep</b>—Lokacija je filtrirana jer je dubina očitavanja preniska</li> <li>• <b>low_frac_info_reads</b>—Lokacija je filtrirana jer je udio informativnih očitavanja ispod granične vrijednosti</li> <li>• <b>low_normalna_depth</b>—Lokacija je filtrirana jer je normalna dubina očitavanja uzorka preniska</li> <li>• <b>long_indel</b>—Lokacija je filtrirana jer je duljina indela preduga</li> <li>• <b>mapping_quality</b>—Lokacija je filtrirana jer srednja kvaliteta mapiranja alt očitavanja na tom lokusu ne zadovoljava granične vrijednosti</li> <li>• <b>multiallelic</b>—Lokacija je filtrirana jer više od dva alt alela prolaze tumorski LOD</li> <li>• <b>non_homref_normal</b>—Lokacija je filtrirana jer genotip normalnog uzorka nije homozigotna referenca</li> <li>• <b>no_reliable_supporting_read</b>—Lokacija je filtrirana jer ne postoji pouzdano prateće somatsko očitavanje</li> <li>• <b>panel_of_normals</b>—primijećeno u najmanje jednom uzorku u panelu normalne vcf datoteke</li> <li>• <b>read_position</b>—Lokacija je filtrirana jer je medijan udaljenosti između početka/kraja očitavanja i ovog lokusa ispod granične vrijednosti</li> <li>• <b>RMxNRepeatRegion</b>—Lokacija je filtrirana jer je cijeli ili dio alelnih varijanti ponavljen u referenci</li> <li>• <b>strand_artifact</b>—Lokacija je filtrirana zbog jake pristranosti niza</li> <li>• <b>str_contraction</b>—Lokacija je filtrirana zbog sumnje na pogrešku PCR-a gdje je alt alel jedna ponovljena jedinica manja od referentne</li> <li>• <b>too_few_supporting_reads</b>—Lokacija je filtrirana jer ima premalo popratnih očitavanja u uzorku tumora</li> </ul>

Zaglavlje	Opis
FILTAR (nastavak)	<ul style="list-style-type: none"> <li>• <b>weak_evidence</b>—Rezultat somatske varijante ne zadovoljava granične vrijednosti</li> </ul> <p>Mogući FILTERA unosi za somatski tijek rada uključuju:</p> <ul style="list-style-type: none"> <li>• <b>base_quality</b>—Lokacija je filtrirana jer srednja osnovna kvaliteta alt očitavanja na tom lokusu ne zadovoljava granične vrijednosti</li> <li>• <b>filtered_reads</b>—Lokacija je filtrirana jer je prevelik dio očitavanja filtriran</li> <li>• <b>fragment_length</b>—Lokacija je filtrirana jer apsolutna razlika između srednje duljine fragmenta alt očitavanja i srednje duljine fragmenta ref očitavanja na tom lokusu prelazi granične vrijednosti</li> <li>• <b>low_deep</b>—Lokacija je filtrirana jer je dubina očitavanja preniska</li> <li>• <b>low_frac_info_reads</b>—Lokacija je filtrirana jer je udio informativnih očitavanja ispod granične vrijednosti</li> <li>• <b>low_normalna_depth</b>—Lokacija je filtrirana jer je normalna dubina očitavanja uzorka preniska</li> <li>• <b>long_indel</b>—Lokacija je filtrirana jer je duljina indela preduga</li> <li>• <b>mapping_quality</b>—Lokacija je filtrirana jer srednja kvaliteta mapiranja alt očitavanja na tom lokusu ne zadovoljava granične vrijednosti</li> <li>• <b>multiallelic</b>—Lokacija je filtrirana jer više od dva alt alela prolaze tumorski LOD</li> <li>• <b>non_homref_normal</b>—Lokacija je filtrirana jer genotip normalnog uzorka nije homozigotna referenca</li> <li>• <b>no_reliable_supporting_read</b>—Lokacija je filtrirana jer ne postoji pouzdano prateće somatsko očitavanje</li> <li>• <b>panel_of_normals</b>—primijećeno u najmanje jednom uzorku u panelu normalne vcf datoteke</li> <li>• <b>read_position</b>—Lokacija je filtrirana jer je medijan udaljenosti između početka/kraja očitavanja i ovog lokusa ispod granične vrijednosti</li> <li>• <b>RMxNRepeatRegion</b>—Lokacija je filtrirana jer je cijeli ili dio alelnih varijanti ponavljen u referenci</li> <li>• <b>strand_artifact</b>—Lokacija je filtrirana zbog jake pristranosti niza</li> <li>• <b>str_contraction</b>—Lokacija je filtrirana zbog sumnje na pogrešku PCR-a gdje je alt alel jedna ponovljena jedinica manja od referentne</li> <li>• <b>too_few_supporting_reads</b>—Lokacija je filtrirana jer ima premalo popratnih očitavanja u uzorku tumora</li> <li>• <b>weak_evidence</b>—Rezultat somatske varijante ne zadovoljava granične vrijednosti</li> <li>• <b>systematic_noise</b>—Lokacija je filtrirana na temelju dokaza sustavne buke u normalnim uvjetima</li> </ul>

Zaglavlje	Opis
INFO	<p>Mogući INFO unosi za tijek rada sa spolnim stanicama uključuju:</p> <ul style="list-style-type: none"> <li>• <b>AC</b>—Broj alela u genotipima za svaki ALT alel, istim redoslijedom kojim su navedeni.</li> <li>• <b>AF</b>—Frekvencija alela za svaki ALT alel, istim redoslijedom kojim su navedeni.</li> <li>• <b>AN</b>—Ukupan broj alela u očitanim genotipima.</li> <li>• <b>DB</b>—dbSNP članstvo.</li> <li>• <b>FS</b>—Phred bodovanje p-vrijednosti pomoću Fisherovog egzaktnog testa za otkrivanje pristranosti niza.</li> <li>• <b>QD</b>—Pouzdanost varijante / kvaliteta prema dubini.</li> <li>• <b>R2_5P_bias</b>—Ocjena na temelju sklonosti niza i udaljenosti od 5 glavnih krajeva.</li> <li>• <b>SOR</b>—Simetrični omjer vjerojatnosti 2x2 tablice nepredviđenosti za otkrivanje pristranosti niza.</li> <li>• <b>DP</b>—Približna dubina očitavanja (informativna i neinformativna); neka očitavanja možda su filtrirana na temelju mapq rezultata itd.</li> <li>• <b>END</b>—Zaustavni položaj intervala.</li> <li>• <b>FractionInformativeReads</b>—Dio informativnih očitavanja od ukupnog broja očitavanja.</li> <li>• <b>MQ</b>—RMS kvaliteta mapiranja.</li> <li>• <b>MQRankSum</b>—Z-rezultat iz Wilcoxonovog testa zbroja rangova Alt u odnosu na Ref očitavanje kvalitete mapiranja.</li> <li>• <b>ReadPosRankSum</b>—Z-rezultat iz Wilcoxonovog testa zbroja rangova Alt u odnosu na Ref pristranost očitavog položaja.</li> <li>• <b>SOMATIC</b>—Najmanje jedna varijanta na tom položaju je somatska.</li> </ul> <p>Mogući INFO unosi za somatski tijek rada uključuju:</p> <ul style="list-style-type: none"> <li>• <b>DP</b>—Približna dubina očitavanja (informativna i neinformativna); neka očitavanja možda su filtrirana na temelju mapq rezultata itd.</li> <li>• <b>END</b>—Zaustavni položaj intervala.</li> <li>• <b>FractionInformativeReads</b>—Dio informativnih očitavanja od ukupnog broja očitavanja.</li> <li>• <b>MQ</b>—RMS kvaliteta mapiranja.</li> <li>• <b>MQRankSum</b>—Z-rezultat iz Wilcoxonovog testa zbroja rangova Alt u odnosu na Ref očitavanje kvalitete mapiranja.</li> <li>• <b>ReadPosRankSum</b>—Z-rezultat iz Wilcoxonovog testa zbroja rangova Alt u odnosu na Ref pristranost očitavog položaja.</li> <li>• <b>AQ</b>—Rezultat sustavne buke.</li> <li>• <b>hotspot</b>—Poznato somatsko mjesto, koristi se za povećanje pouzdanosti očitavanja.</li> <li>• <b>SOMATIC</b>—Najmanje jedna varijanta na tom položaju je somatska.</li> </ul>



Zaglavlje	Opis
FORMAT	<p>U stupcu Format navedena su polja, odvojena dvotočkama. Npr. GT:GQ.</p> <p>Dostupna polja za tijek rada sa spolnim stanicama uključuju:</p> <ul style="list-style-type: none"> <li>• <b>AD</b>—Dubine alela (koje sadrže samo informativne očitavanja od ukupnog broja očitavanja) za ref i alt alele navedenim redoslijedom.</li> <li>• <b>AF</b>—Fracije alela za alt alele navedenim redoslijedom.</li> <li>• <b>DP</b>—približna dubina očitavanja (filtriraju se očitavanja s MQ=255 ili lošim parovima).</li> <li>• <b>F1R2</b>—Broj očitavanja u parnom usmjerenju F1R2 koja podržavaju svaki alel.</li> <li>• <b>F2R1</b>—Broj očitavanja u parnom usmjerenju F2R1 koja podržavaju svaki alel.</li> <li>• <b>GP</b>—Posteriorne vjerojatnosti prema Phred bodovanju za genotipove prema definiciji u VCF specifikaciji.</li> <li>• <b>GQ</b>—Kvaliteta genotipa.</li> <li>• <b>GT</b>—Genotip. 0 odgovara referentnoj bazi, 1 odgovara prvom unosu u stupac ALT itd. Kosa crta (/) ukazuje da nema dostupnih informacija o fazama.</li> <li>• <b>MB</b>—Statistika komponente po uzorku za otkrivanje sklonosti niza.</li> <li>• <b>PL</b>—Normalizirano, Phred bodovanje vjerojatnosti za genotipove kako je definirano u VCF specifikaciji.</li> <li>• <b>PRI</b>—Prethodne vjerojatnosti prema Phred bodovanju za genotipove.</li> <li>• <b>PS</b>—Informacije o ID-u fizičkog faziranja, pri čemu svaki jedinstveni ID unutar određenog uzorka (ali ne između uzoraka) povezuje zapise unutar grupe za faziranje.</li> <li>• <b>SB</b>—Statistika komponenti po uzorku koja se sastoji od Fisherovog egzaktnog testa za otkrivanje pristranosti niza.</li> <li>• <b>SQ</b>—Somatska kvaliteta.</li> </ul> <p>Dostupna polja somatskog tijeka rada uključuju:</p> <ul style="list-style-type: none"> <li>• <b>AD</b>—Dubine alela (koje sadrže samo informativne očitavanja od ukupnog broja očitavanja) za ref i alt alele navedenim redoslijedom.</li> <li>• <b>AF</b>—Fracije alela za alt alele navedenim redoslijedom.</li> <li>• <b>DP</b>—približna dubina očitavanja (filtriraju se očitavanja s MQ=255 ili lošim parovima).</li> <li>• <b>F1R2</b>—Broj očitavanja u parnom usmjerenju F1R2 koja podržavaju svaki alel.</li> <li>• <b>F2R1</b>—Broj očitavanja u parnom usmjerenju F2R1 koja podržavaju svaki alel.</li> <li>• <b>GT</b>—Genotip. 0 odgovara referentnoj bazi, 1 odgovara prvom unosu u stupac ALT itd. Kosa crta (/) ukazuje da nema dostupnih informacija o fazama.</li> <li>• <b>MB</b>—Statistika komponente po uzorku za otkrivanje sklonosti niza.</li> <li>• <b>PS</b>—Informacije o ID-u fizičkog faziranja, pri čemu svaki jedinstveni ID unutar određenog uzorka (ali ne između uzoraka) povezuje zapise unutar grupe za faziranje.</li> <li>• <b>SB</b>—Statistika komponenti po uzorku koja se sastoji od Fisherovog egzaktnog testa za otkrivanje pristranosti niza.</li> <li>• <b>SQ</b>—Somatska kvaliteta.</li> </ul>
UZORAK	Stupac uzorka sadrži vrijednosti navedene u stupcu FORMAT.

## Genomske VCF datoteke

Genomske VCF (\*.gvcf.gz) datoteke slijede niz dogovorenih pravila za predstavljanje svih mjesta u genomu u relativno kompaktnom obliku. gVCF datoteke obuhvaćaju sva mjesta unutar područja od interesa u jednoj datoteci za svaki uzorak. gVCF datoteka pokazuje neočitavanja na položajima koji ne zadovoljavaju sve filtre. Oznaka genotipa (GT) ./. ukazuje na neočitavanje.

## Prikaz rezultata analize

Obrade koje su trenutačno u tijeku prikazuju se na kartici Active (Aktivno). Dovořeni ciklusi obrade prikazuju se na kartici Completed (Dovořeno). Pogledajte [Dokumentacija proizvoda NovaSeq 6000Dx \(broj dokumenta 200010105\)](#) za više informacija o prikazu rezultata.

# Tehnička pomoć

Ako vam je potrebna tehnička pomoć, obratite se Illumina Tehnička podrška.

**Web-mjesto:** [www.illumina.com](http://www.illumina.com)  
**Adresa e-pošte:** [techsupport@illumina.com](mailto:techsupport@illumina.com)

## Illumina Telefonski brojevi službe za tehničku podršku

Regija	Besplatni telefon	Međunarodni
Australija	+61 1800 775 688	
Austrija	+43 800 006249	+43 1 9286540
Belgija	+32 800 77 160	+32 3 400 29 73
Kanada	+1 800 809 4566	
Kina		+86 400 066 5835
Danska	+45 80 82 01 83	+45 89 87 11 56
Finska	+358 800 918 363	+358 9 7479 0110
Francuska	+33 8 05 10 21 93	+33 1 70 77 04 46
Njemačka	+49 800 101 4940	+49 89 3803 5677
Hong Kong, Kina	+852 800 960 230	
Indija	+91 8006500375	
Indonezija		0078036510048
Irska	+353 1800 936608	+353 1 695 0506
Italija	+39 800 985513	+39 236003759
Japan	+81 0800 111 5011	
Malezija	+60 1800 80 6789	
Nizozemska	+31 800 022 2493	+31 20 713 2960
Novi Zeland	+64 800 451 650	
Norveška	+47 800 16 836	+47 21 93 96 93
Filipini	+63 180016510798	
Singapur	1 800 5792 745	
Južna Koreja	+82 80 234 5300	

Regija	Besplatni telefon	Međunarodni
Španjolska	+34 800 300 143	+34 911 899 417
Švedska	+46 2 00883979	+46 8 50619671
Švicarska	+41 800 200 442	+41 56 580 00 00
Tajvan, Kina	+886 8 06651752	
Tajland	+66 1800 011 304	
Velika Britanija	+44 800 012 6019	+44 20 7305 7197
Sjedinjene Američke Države	+1 800 809 4566	+1 858 202 4566
Vijetnam	+84 1206 5263	

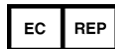
**Sigurnosno-tehnički listovi (SDS-ovi)**—Dostupni su na web-mjestu tvrtke Illumina na adresi [support.illumina.com/sds.html](https://support.illumina.com/sds.html).

**Dokumentacija o proizvodima**—Dostupna je za preuzimanje na web-mjestu [support.illumina.com](https://support.illumina.com).



Illumina  
5200 Illumina Way  
San Diego, California 92122 SAD  
+1.800.809.ILMN (4566)  
+1.858.202.4566 (izvan Sjeverne Amerike)  
techsupport@illumina.com  
www.illumina.com

CE



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

**Australski sponzor**

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

ZA IN VITRO DIJAGNOSTIKU

© 2022. Illumina, Inc. Sva prava pridržana.

**illumina**<sup>®</sup>