

Modul za analizu Local Run Manager Germline Variant

Vodič za tijek rada za NextSeq 550Dx

ZA IN VITRO DIJAGNOSTIKU

Pregled	3
Unos podataka o obradi	3
Metode analize	5
Prikaz podataka o obradi i uzorku	7
Izvješće o analizi	7
Izlazne datoteke analize	8
Povijest revizija	15
Tehnička pomoć	16



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 proizvodima opisanima u njemu. 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 opisanih 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).

© 2021. Illumina, Inc. Sva prava pridržana.

Svi su žigovi vlasništvo tvrtke Illumina, Inc. i njezinih vlasnika. Konkretno informacije o žigovima potražite na adresi www.illumina.com/company/legal.html.

Pregled

Local Run Manager Germline Variant Module namijenjen je upotrebi s analizom Illumina TruSeq Custom Amplicon Kit Dx i instrumentom NextSeq 550Dx. Kad se upotrebljava sa softverom Germline Variant Module, analiza je namijenjena pripremi biblioteka korištenih za sekvenciranje DNA iz uzoraka periferne pune krvi.

Modul za analizu pregledava varijante u kratkim područjima amplificirane DNA ili amplikonima. Fokusrano sekvenciranje amplikona omogućuje visoku razinu pokrivenosti određenih područja. Pogledajte informativni pregled za *TruSeq Custom Amplicon Kit Dx* (broj dokumenta: 1000000029772).

Modul za analizu Germline Variant upotrebljava potrošni materijal za sekvenciranje u 300 ciklusa. Da biste saznali više, pročitajte informativne preglede za *NextSeq 550Dx High Output Reagent Kit v2* ili *NextSeq 550Dx High Output Reagent Kit v2.5*.

O ovom vodiču

U ovom vodiču navedene su upute za postavljanje parametara obrade za sekvenciranje i analizu putem modula za analizu Germline Variant. Da biste saznali više o nadzornoj ploči softvera Local Run Manager i postavkama sustava, pročitajte *Referentni priručnik za instrument NextSeq 550Dx* (broj dokumenta: 1000000009513).

Prikaz softvera Local Run Manager

Sučelje softvera Local Run Manager prikazuje se u softveru NextSeq 550Dx Operating Software (NOS) ili putem web-preglednika. Podržan je web-preglednik Chromium.



NAPOMENA

Ako upotrebljavate preglednik koji nije podržan, preuzmite podržani preglednik kad se to od vas zatraži u poruci "Confirm Unsupported Browser" (Potvrda nepodržanog preglednika). Odaberite "**here**" (ovdje) da biste preuzeli podržanu verziju Chromioma.

Prikaz na monitoru instrumenta

- 1 Da biste na monitoru instrumenta prikazali sučelje softvera Local Run Manager, odaberite neku od sljedećih mogućnosti:
 - ▶ Na početnom zaslonu NOS-a odaberite **Local Run Manager**.
Kliknite X u gornjem desnom kutu da biste se vratili na NOS kad završite.
 - ▶ Odaberite ikonu Minimize NOS (Minimiziraj NOS), otvorite web-preglednik Chromium koji je na instrumentu i u traku adrese unesite **http://localhost**.
Samo korisnici administratori mogu minimizirati NOS.

Prikaz na umreženom računalu

- 1 Otvorite web-preglednik Chromium na računalu koje ima pristup istoj mreži kao instrument i povežite se s instrumentom koristeći se IP adresom ili nazivom instrumenta. Na primjer, **http://myinstrument**.

Unos podataka o obradi

Postavljanje parametara

- 1 Prijavite se u Local Run Manager.
- 2 Odaberite **Create Run** (Stvori obradu) pa **Germline Variant**.

- Unesite naziv obrade po kojem ćete razlikovati obradu od sekvenciranja preko analize. Pritom upotrijebite alfanumeričke znakove, razmake, podvlake ili crtice.
- [neobavezno]** Unesite opis obrade da biste je lakše razlikovali. Pritom upotrijebite alfanumeričke znakove, razmake, podvlake ili crtice.
- Na padajućem popisu odaberite broj uzoraka i skup indeksa. Prilikom odabira uzmite u obzir sljedeće informacije:
 - ▶ Padajući popis sadrži brojeve uzoraka sa skupom indeksa. Primjerice, 24-Set 1 znači da će se testirati 24 uzorka s indeksima iz skupa indeksa 1.
 - ▶ Brojevi skupova indeksa odnose se na različite skupove i5 indeksa. Skup 1 i skup 2 omogućuju raznolikost indeksa. Ponuđena su dva skupa indeksa kako bi se spriječilo iscrpljivanje jednog skupa.
 - ▶ Odaberite broj uzoraka koji je najbliži broju uzoraka koji testirate. Ako točan broj uzoraka nije na popisu, odaberite broj koji je najbliži ali manji od broja koji testirate. Primjerice, ako želite testirati 18 uzoraka, odaberite 16 uzoraka.
 - ▶ Jažice s uzorcima i kombinacije indeksa koji zadovoljavaju preduvjete raznolikosti indeksa istaknuti su zelenom bojom. Ako odaberete druge jažice i kombinacije indeksa, kad spremite obradu primit ćete obavijest ako preduvjete raznolikosti indeksa nisu zadovoljeni.

Uvoz datoteka manifesta za obradu

- Pazite da manifesti koje želite uvesti budu dostupni na pristupačnoj mrežnoj lokaciji ili na USB pogonu.
- Odaberite **Import Manifests** (Uvezi manifeste).
- Idite na datoteku manifesta i odaberite manifest koji želite dodati.



NAPOMENA

Da biste datoteke manifesta učinili dostupnima svim obradama pomoću modula za analizu Germline Variant, dodajte manifeste putem značajke Module Settings (Postavke modula). Ta značajka zahtijeva dozvole za korisničku razinu administratora. Da biste saznali više, pročitajte *Referentni priručnik za instrument NextSeq 550Dx* (broj dokumenta: 1000000009513).

Navođenje uzoraka za obradu

Navedite uzorke za obradu pomoću jedne od mogućnosti i sljedećih smjernica.


- ▶ **Ručni unos uzoraka** – upotrijebite praznu tablicu na zaslonu Create Run (Stvori obradu).
- ▶ **Uvoz uzoraka** – otvorite vanjsku datoteku u formatu vrijednosti odvojenih zarezom (*.csv). Moguće je preuzeti predložak na zaslonu Create Run (Stvori obradu).

Kad popunite tablicu s uzorcima, možete izvesti podatke o uzorcima u vanjsku datoteku. Upotrebite tu datoteku kao referencu prilikom pripreme biblioteka ili je uvezite za neku drugu obradu.


Ručni unos uzoraka

- U polje Sample ID (ID uzorka) unesite jedinstveni ID uzorka. Pritom upotrebite alfanumeričke znakove, crtice ili podvlake.
- [neobavezno]** Za pozitivne ili negativne kontrolne uzorke desnom tipkom miša kliknite i odaberite vrstu kontrole.
- [neobavezno]** U polje Sample Description (Opis uzorka) unesite opis uzorka. Pritom upotrebite alfanumeričke znakove, crtice ili podvlake.

Opisi uzoraka povezuju se s nazivom uzoraka. Opisi uzoraka prepisuju se ako se isti naziv uzorka ponovno upotrebljava u kasnijoj obradi.

- 4 Odaberite prilagodnik Indeks 1 na padajućem popisu Index 1 (i7).
Kad upotrebljavate predložene jažice za uzorke, softver automatski popunjava prilagodnike indeksa i7 i i5 koji zadovoljavaju preduvjete različitosti za indekse. Ako točan broj uzoraka koji testirate nije na popisu, pripazite da odaberete prilagodnike indeksa za dodatne jažice. Ako trebate odabrati indekse za dodatne jažice ili ne upotrebljavate predložene kombinacije prilagodnika indeksa, prije nego što odaberete indekse svakako pročitajte *Prepoznavanje baza i raznolikost indeksa na stranici 14*.
- 5 Odaberite prilagodnik Indeks 2 na padajućem popisu Index 2 (i5).
- 6 Odaberite datoteku manifesta na padajućem popisu manifesta.
- 7 Odaberite mogućnost za prikaz, ispis ili spremanje rasporeda pločice kao referencu za pripremu biblioteka:
 - ▶ Odaberite ikonu  **Print** (Ispis) da biste prikazali raspored pločice. Odaberite **Print** (Ispis) da biste ispisali raspored pločice.
 - ▶ Odaberite **Export** (Izvezi) da biste izvezli podatke o uzorku u vanjsku datoteku.Provjerite jesu li podaci iz manifesta i oni o uzorku točni. Netočni podaci mogu utjecati na rezultate.
- 8 Odaberite **Save Run** (Spremi obradu).

Uvoz uzoraka

- 1 Odaberite **Import Samples** (Uvezi uzorke) i idite na mjesto na kojem se nalazi datoteka s podacima o uzorku. Možete uvesti dvije vrste datoteka.
 - ▶ Na zaslону Create Run (Stvori obradu) odaberite **Template** (Predložak) da biste napravili novi raspored pločice. Datoteka predloška sadrži točna zaglavlja stupaca za uvoz. U svaki stupac unesite podatke o uzorku za uzorke u obradi. Izbrišite primjere podataka u člancima koji se ne upotrebljavaju, a zatim spremite datoteku.
 - ▶ Upotrijebite datoteku s podacima o uzorcima koja je izvezena iz modula Germline Variant pomoću značajke za izvoz.
- 2 Odaberite ikonu  **Print** (Ispis) da biste prikazali raspored pločice.
- 3 Odaberite **Print** (Ispis) da biste ispisali raspored pločice kao referencu za pripremu biblioteka.
- 4 **[neobavezno]** Odaberite **Export** (Izvezi) da biste izvezli podatke o uzorku u vanjsku datoteku. Provjerite jesu li podaci iz manifesta i oni o uzorku točni. Netočni podaci mogu utjecati na rezultate.
- 5 Odaberite **Save Run** (Spremi obradu).

Uređivanje obrade

Upute za uređivanje podataka o obradi prije sekvenciranja potražite u *Referentnom vodiču za instrument NextSeq 550Dx (broj dokumenta: 1000000009513)*.

Metode analize

Modul za analizu Germline Variant izvodi sljedeće korake analize, a zatim zapisuje izlazne datoteke analize u mapuAlignment (Usklađivanje).

- ▶ Čitanja indeksa demultipleksiranja
- ▶ Generiranje FASTQ datoteka

- ▶ Usklađivanje s referencom
- ▶ Identifikacija varijanti

Demultipleksiranje

Demultipleksiranje uspoređuje svaku sekvencu čitanja indeksa s navedenim sekvencama indeksa za obradu. U tom se koraku ne gledaju vrijednosti kvalitete.

Čitanja 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 klasterne koji nisu dodijeljeni uzorku.
- ▶ Klasteri se dodjeljuju uzorku kad se sekvenca indeksa točno podudara ili kad se nađe maksimalno jedno nepodudaranje po čitanju indeksa.

Generiranje FASTQ datoteke

Nakon demultipleksiranja softver generira sporedne datoteke analize u formatu FASTQ, a to je tekstni format korišten za predstavljanje sekvenci. FASTQ datoteke sadrže očitavanja za svaki uzorak i povezano bodovanje kvalitete. Klasteri koji nisu zadovoljili filter izostavljaju se.

Svaka FASTQ datoteka sadrži očitavanja samo za jedan uzorak, a naziv tog uzorka uvršten je u naziv FASTQ datoteke. FASTQ datoteke primarne su ulazne datoteke za usklađivanje. Po uzorku se generira osam FASTQ datoteka – četiri iz 1. čitanja i četiri iz 2. čitanja.

Usklađivanje

Tijekom koraka usklađivanja stupnjeviti Smith-Watermanov algoritam usklađuje klasterne iz svakog uzorka sa sekvencama amplikona navedenim u datoteci manifesta.

Stupnjeviti Smith-Watermanov algoritam izvodi poluglobalno usklađivanje sekvenci radi određivanja sličnih područja između dviju sekvenci. Umjesto uspoređivanja ukupne sekvence, Smith-Watermanov algoritam uspoređuje segmente svih mogućih dužina.

Svako čitanje s uparenim krajevima procjenjuje se s obzirom na usklađenost s relevantnim probnim sekvencama za to čitanje.

- ▶ Prvo čitanje procjenjuje se u odnosu na reverzni komplement silaznih oligonukleotida specifičnih za lokus (Downstream Locus-Specific Oligos, DLSO).
- ▶ Drugo čitanje procjenjuje se u odnosu na ulazne oligonukleotide specifične za lokus (Upstream Locus-Specific Oligos, ULSO).
- ▶ Ako se početak čitanja podudara s probnom sekvencom uz ne više od tri razlike (nepodudaranja ili pomaci zbog vodećih indela), puna dužina čitanja usklađuje se s amplikonskim ciljem za tu sekvencu.
- ▶ Indeli u DLSO-u i ULSO-u ne promatraju se zbog kemije analize.

Usklađivanja se filtriraju iz rezultata usklađivanja na temelju omjera nepodudaranja u području interesa ili potpunog amplikona, ovisno o dužini amplikona. Filtrirana usklađivanja zapisuju se u datoteke usklađivanja kao neusklađena i ne upotrebljavaju se u prepoznavanju varijante.

Prepoznavanje varijante

Prepoznavatelj varijante Pisces koji je razvila tvrtka Illumina prepoznaje varijante prisutne u uzorku DNA.

Prepoznavatelj varijanti Pisces prepoznaje SNV-ove, MNV-ove i male indele u 3 koraka:



- ▶ uzima u obzir svaki položaj u referentnom genomu zasebno

- ▶ broji baze na odgovarajućem položaju i traži očitavanja koja se preklapaju na položaju
- ▶ računa ocjenu varijante kojim se mjeri kvaliteta prepoznavanja na temelju Poissonova modela; varijante s ocjenom kvalitete manjom od Q20 izbacuju se.

Ako varijanta zadovolji sve filtre, ona se u VCF-u označava s PASS (Zadovoljava).

Da biste saznali više, pročitajte github.com/Illumina/Pisces/wiki.

Prikaz podataka o obradi i uzorku

- 1 Na nadzornoj ploči programa Local Run Manager kliknite naziv obrade.
- 2 Na kartici Run Overview (Pregled obrade) pregledajte metriku obrade sekvenciranja.
- 3 **[neobavezno]** Kliknite ikonu **Copy to Clipboard**  (Kopiraj u međuspremnik) da biste kopirali stazu izlazne mape obrade.
- 4 Kliknite karticu Sequencing Information (Informacije o sekvenciranju) da biste pregledali parametre obrade i informacije o potrošnom materijalu.
- 5 Kliknite karticu Samples and Results (Uzorci i rezultati) da biste prikazali lokaciju izvješća nakon analize.
 - ▶ Ako se analiza ponovila, proširite padajući popis Select Analysis (Odabir analize) i odaberite odgovarajuću analizu.
- 6 Kliknite ikonu **Copy to Clipboard**  (Kopiraj u međuspremnik) da biste kopirali stazu mape Analiza.

Da biste saznali više o karticama Pregled obrade i Informacije o sekvenciranju te kako ponovno staviti analizu u red čekanja, pročitajte *Referentni vodič za instrument NextSeq 550Dx (broj dokumenta: 1000000009513)*.

Izvješće o analizi

Rezultati analize sažeti su na karticama Samples (Uzorci) i Results (Rezultati) te kao zbirno izvješće u mapi Alignment (Usklađivanje). Izvješće za svaki uzorak dostupno je i u formatu datoteke PDF za svaki uzorak.

Podaci na karticama Samples (Uzorci) i Results (Rezultati)

- 1 Kliknite neki uzorak na popisu da biste vidjeli izvješće o uzorku.

Tbl. 1 Podaci o obradi i uzorku

Zaglavlje stupca	Opis
Run Status (Status obrade)	Govori je li obrada sekvenciranja uspjela ili nije uspjela.
Total Yield (GB) (Ukupan prinos (GB))	Broj baza prepoznatih u obradi sekvenciranja. Pokazuje prag prolaznosti i status uspješnosti ili neuspješnosti.
% ≥ Q30	Postotak čitanja u obradi sekvenciranja s ocjenom kvalitete 30 (Q30) ili većom. Pokazuje prag prolaznosti i status uspješnosti ili neuspješnosti.
ID uzorka	ID uzorka naveden prilikom stvaranja obrade.
Total PF Reads (Ukupan broj čitanja sa statusom PF)	Ukupan broj čitanja koja zadovoljavaju filter.
Read 1% ≥ Q30 (% ≥ Q30 za 1. čitanje)	Postotak čitanja u 1. čitanju s ocjenom kvalitete za uzorak 30 (Q30) ili većom.

Zaglavlje stupca	Opis
Read 2% \geq Q30 (% \geq Q30 za 2. čitanje)	Postotak čitanja u 2. čitanju s ocjenom kvalitete za uzorak 30 (Q30) ili većom.
Autosome Call Rate (Omjer prepoznavanja autosoma)	Broj genomičkih položaja u autosomima (od 1. do 22. kromosoma) koji zadovoljavaju unaprijed definiranu vrijednost granice pouzdanosti podijeljen s ukupnim brojem ispitivanih autosomnih genomičkih položaja. Omjer prepoznavanja opisan je za uzorak i izražen kao postotak koji se računa kao 1 minus (broj autosomnih položaja s nepotpunim prepoznavanjem podijeljen s ukupnim brojem sekvenciranih autosomnih položaja).

Tbl. 2 Podaci izvješća o uzorku

Zaglavlje stupca	Opis
Sample (Uzorak)	ID uzorka naveden prilikom stvaranja obrade.
Report Date (Datum izvješća)	Datum kad je izvješće generirano.
Podaci o uzorku	ID uzorka koji je naveden prilikom stvaranja obrade, ukupan broj čitanja u uzorku koja su zadovoljila filter, postotak čitanja za uzorak s ocjenom kvalitete 30 (Q30) ili većom i omjer prepoznavanja autosoma.
Amplicon Summary (Sažetak amplikona)	Ukupan broj sekvenciranih regija amplikona i ukupna dužina u parovima baza sekvenciranih amplikona u ciljnim područjima za uzorak i datoteku manifesta. Datoteka manifesta određuje referentni genom i ciljna referentna područja korištena u koraku usklađivanja.
Read Level Statistics (Statistički podaci na razini očitavanja)	Broj i postotak čitanja za uzorak koji pokrivaju svaki položaj u referenci, za 1. čitanje i 2. čitanje.
Variants Summary (Sažetak varijanti)	Broj SNV-ova, umetanja i brisanja prepoznatih za uzorak koji je prošao predložene vrijednosti za određivanje jesu li rezultati kvalitete u prihvatljivom rasponu.
Coverage Summary (Sažetak pokrivenosti)	Ukupan broj usklađenih baza podijeljen prema veličini ciljnog područja i postotak ampliconskih područja s vrijednostima pokrivenosti većim od niskog praga pokrivenosti od 0,2 * srednja pokrivenost amplikona za uzorak.
Coverage Plots (Skice pokrivenosti)	Skica pokrivenosti prema području amplikona prikazuje pokrivenost u područjima amplikona za uzorak. Područja s vrijednostima pokrivenosti manjima od praga pokrivenosti istaknuta su crvenom bojom. Prosjek svih vrijednosti naznačen je narančastom crtom.
Software Versions (Verzije softvera)	Verzije programa pomoću kojih je uzorak sekvenciran. Uključuje softver NextSeq 550Dx Operating Software (NOS), softver Local Run Manager, softver RTA i verziju Germline Variant Module.

Izlazne datoteke analize

Sljedeće se izlazne datoteke analize generiraju za modul za analizu Germline Variant i donose rezultate analize za usklađivanje i prepoznavanje varijanti. Izlazne datoteke analize nalaze se u mapi Alignment (Usklađivanje).

Naziv datoteke	Opis
Demultipleksiranje (*.txt)	Sporedne datoteke koje sadrže sažete rezultate demultipleksiranja.
FASTQ (*.fastq.gz)	Sporedne datoteke koje sadrže prepoznavanja baza za koja postoji ocjena kvalitete. FASTQ datoteke primarne su ulazne datoteke u koraku usklađivanja.
Datoteke usklađivanja u formatu BAM (*.bam)	Sadrže usklađena očitavanja za određeni uzorak.

Naziv datoteke	Opis
Datoteke prepoznavanja varijanti u genomskom formatu VCF (*.genome.vcf.gz)	Sadrži genotip za svaki položaj, prepoznat kao varijanta ili prepoznat kao referenca.
Datoteke prepoznavanja varijanti u VCF formatu (*.vcf.gz)	Sadrži sve varijante prepoznate u cijelom ciljnom području.
AmpliconCoverage_M1.tsv	Sadrži informacije o pokrivenosti prema amplikonu po uzorku za svaki navedeni manifest. M# predstavlja broj manifesta.

Format datoteke demultipleksiranja

Postupak demultipleksiranja čita sekvencu indeksa priloženu svakom klasteru radi određivanja iz kojeg uzorka klaster potječe. Mapiranje između klastera i broja uzorka zapisuje se u datoteku demultipleksiranja (*.demux) za svaku pločicu protočne jedinice.

Format imenovanja datoteka demultipleksiranja je **s_1_X.demux**, pri čemu je X broj pločice.

Datoteke demultipleksiranja počinju zaglavljem:

- ▶ verzija (cjelobrojna vrijednost od 4 bajta), trenutno je to 1
- ▶ broj klastera (cjelobrojna vrijednost od 4 bajta)

Ostatak datoteke sastoji se od brojeva uzoraka iz svakog klastera s pločice.

Kad se korak demultipleksiranja završi, softver generira datoteku demultipleksiranja naziva **DemultiplexSummaryF1L1.txt**.

- ▶ U tom nazivu datoteke **F1** predstavlja broj protočne jedinice.
- ▶ U tom nazivu datoteke **L1** predstavlja broj staze.
- ▶ Rezultati demultipleksiranja u tablici s 1 retkom po pločici i 1 stupcem po uzorku, uključujući uzorak 0.
- ▶ Sekvence u čitanjima indeksa koje se najčešće pojavljuju.

Format datoteke FASTQ

FASTQ je format datoteke utemeljen na tekstu koji sadrži prepoznavanja baza i vrijednosti kvalitete po očitavanju. Svaki zapis sastoji se od 4 retka:

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

Identifikator je ovako oblikovan:

@Instrument:IDObrade:IDProtočneJedinice:Staza:Pločica:X:Y BrojČitanja:ZastavicaFiltra:0:BrojUzorka

Primjer:

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

Format datoteke BAM

BAM datoteka (*.bam) je komprimirana binarna verzija SAM datoteke koja se upotrebljava za predstavljanje usklađenih sekvenci veličine do 128 Mb. Format SAM i BAM detaljno su opisani u dokumentu samtools.github.io/hts-specs/SAMv1.pdf.

BAM datoteke upotrebljavaju oblik imenovanja datoteka **NazivUzorka_S#.bam**, pri čemu je # broj uzorka određen redoslijedom kojim su uzorci navedeni za obradu.

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 čitanja, kvalitetu čitanja, 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 čitanja koja upućuje na broj čitanja 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 očitanjem

BAM datoteke indeksa (*.bam.bai) sadrže indeks odgovarajuće BAM datoteke.

Format datoteke VCF

Variant Call Format (VCF, format prepoznavanja varijante) uobičajen je format datoteke koji je razvila genomička znanstvena zajednica. Sadrži informacije o varijantama pronađenim na određenim položajima u referentnom genomu. VCF datoteke završavaju nastavkom .vcf

Zaglavlje VCF datoteke sastoji se od verzije formata datoteke VCF i verzije prepoznavatelja varijanti 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.

Zaglavlja VCF datoteka

Zaglavlje	Opis
CHROM	Kromosom referentnog genoma. Kromosomi se pojavljuju istim redoslijedom kao u referentnoj datoteci FASTQ.
POS	Jednobazni položaj varijante u referentnom kromosomu. Za SNP-ove taj je položaj referentna baza u varijanti; za indele ili brisanja (delecije) taj je položaj referentna baza netom prije varijante.
ID	Rs broj varijante dobiven iz datoteke dbSNP.txt ako je to primjenljivo. Ako na tom mjestu postoji više rs brojeva, popis je odijeljen točka-zarezima. Ako na tom položaju ne postoji dbSNP unos, primjenjuje se marker za vrijednost koja nedostaje ('.').

Zaglavlje	Opis
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 prepoznavaatelj 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 prepoznavanja ima omjer pogrešaka od 0,1 %. Mnogi prepoznavaatelji varijanti boduju kvalitetu na temelju svojih statističkih modela, koji visoko koreliraju s uočenim omjerom pogrešaka.

Oznake u VCF datoteci

Zaglavlje	Opis
FILTER (Filtar)	Ako je datoteka prošla sve filtre, u stupcu filtra piše PASS (Prolaz). <ul style="list-style-type: none"> • LowDP – primijenjeno na mjesta s dubinom pokrivenosti manjom od 150x. Kod položaja amplicona pokrivenih naprednim i reverznim čitanjem to je ekvivalentno 300 preklapajućih čitanja s uparenim krajevima. • q20 – bodovanje kvalitete < 20. • MultiAllelicSite – varijanta ne odgovara diploidnom modelu. • R5x9 – broj ponavljanja (dužine od 1 do 5 bp) susjednih varijanti ≥ 9. • SB – pristranost lanca veća je od zadanog praga.
INFO	Mogući unosi u stupcu INFO uključuju sljedeće: <ul style="list-style-type: none"> • AC – broj alela u genotipima za svaki ALT alel, istim redoslijedom kojim su navedeni. • AF – frekvencija alela za svaki ALT alel, istim redoslijedom kojim su navedeni. • AN – ukupan broj alela u prepoznatim genotipima. • CD – zastavica koja označava da se SNP pojavljuje u području kodiranja barem 1 unosa RefGene. • DP – dubina (broj prepoznavanja baza podudarnih s položajem i korištenih u prepoznavanju varijanti). • Exon – popis očitavanja područja egzona, odvojenih zarezom, iz datoteke RefGene. • FC – funkcionalna posljedica. • GI – popis očitavanja ID-a gena, odvojenih zarezom, iz datoteke RefGene. • QD – pouzdanost varijante/kvaliteta prema dubini. • TI – popis očitavanja ID-a transkripta, odvojenih zarezom, iz datoteke RefGene.
FORMAT	U stupcu Format navedena su polja, odvojena dvotočkama. Npr. GT:GQ. Navedeni popis polja ovisi o korištenom prepoznavaatelju varijanti. Dostupna su sljedeća polja: <ul style="list-style-type: none"> • AD – unos oblika X,Y, pri čemu je X broj referentnih prepoznavanja, a Y je broj alternativnih prepoznavanja. • DP – približna dubina čitanja; filtriraju se očitavanja s MQ=255 ili lošim parovima. • GQ – kvaliteta genotipa. • GT – genotip. 0 odgovara referentnoj bazi, 1 odgovara prvom unosu u stupac ALT itd. Kosa crta (/) ukazuje da nema dostupnih informacija o fazama. • NC – dio baza koje nisu prepoznate ili s kvalitetom prepoznavanja baze manjom od minimalnog praga. • NL – razina šuma; procjena šuma u prepoznavanju baze na tom položaju. • SB – pristranost lanca na tom mjestu. Veće negativne vrijednosti ukazuju na manju pristranost; vrijednosti bliske 0 ukazuju na veću pristranost. • VF – frekvencija varijante; postotak očitavanja koji podržava alternativni alel.
SAMPLE (Uzorak)	Stupac uzorka sadrži vrijednosti navedene u stupcu FORMAT.

Genomske VCF datoteke

Genomske VCF (gVCF) datoteke su VCF v4.1 datoteke koje slijede niz dogovorenih pravila za predstavljanje svih mjesta u genomu u relativno kompaktnom obliku. gVCF (*.genome.vcf.gz) datoteke obuhvaćaju sva mjesta unutar područja od interesa u jednoj datoteci za svaki uzorak.

gVCF datoteka pokazuje neprepoznavanja na položajima koji ne zadovoljavaju sve filtre. Genotipska (GT) oznaka ./. znači neprepoznavanje.

Da biste saznali više, pročitajte sites.google.com/site/gvcftools/home/about-gvcf.

Datoteka pokrivenosti amplikona

Datoteka pokrivenosti amplikona generira se za svaku datoteku manifesta. M# u nazivu datoteke predstavlja broj manifesta.

Svaka datoteka sadrži redak zaglavlja koji obuhvaća ID-ove uzoraka povezane s manifestom. Ta datoteka sadrži sljedeće informacije.

- ▶ Ciljni ID kako je naveden u manifestu.
- ▶ Dubina pokrivenosti očitavanja koja zadovoljavaju filter.

Pomoćne izlazne datoteke

U sljedećim se izlaznim datotekama nalaze pomoćne informacije ili sažeci rezultata obrade i pogrešaka u analizi. Te datoteke nisu nužne za procjenu rezultata analize, ali mogu se upotrebljavati u svrhe otklanjanja poteškoća. Ako nije drugačije navedeno, sve se datoteke nalaze u mapi Alignment (Usklađivanje).

Naziv datoteke	Opis
AnalysisLog.txt	Zapisnik o obradi koji opisuje svaki korak poduzet tijekom analize trenutačne mape za obradu. Ta datoteka ne sadrži poruke o pogreškama. Nalazi se u mapi Alignment (Usklađivanje).
AnalysisError.txt	Zapisnik o obradi u kojem se navode sve pogreške koje su se dogodile tijekom analize. Ta će datoteka biti prazna ako se nije pojavila nijedna pogreška. Nalazi se u mapi Alignment (Usklađivanje).
DemultiplexSummaryF1L1#.txt	Izvešćuje o rezultatima demultipleksiranja u tablici s 1 retkom po pločici i 1 stupcem po uzorku. Broj predstavlja staze 1, 2, 3 ili 4 protočne jedinice. Nalazi se u mapi Alignment (Usklađivanje).
AmpliconRunStatistics.xml	Sadrži sažetak statističkih podataka specifičnih za konkretnu obradu. Nalazi se u mapi Alignment (Usklađivanje).

Mapa Analysis (Analiza)

U mapi analize nalaze se datoteke koje generira softver Local Run Manager.


Odnos između izlazne mape i mape analize može se sažeti ovako:

- ▶ Tijekom sekvenciranja analiza u stvarnom vremenu (Real-Time Analysis, RTA) puni izlaznu mapu datotekama generiranim tijekom analize slike, prepoznavanja baza i bodovanja kvalitete.
- ▶ RTA u stvarnom vremenu kopira datoteke u mapu analize. Kad RTA dodijeli bodovanje kvalitete svakoj bazi iz svakog ciklusa, softver zapisuje datoteku RTAComplete.txt u obje mape.
- ▶ Kad je prisutna datoteka RTAComplete.txt, počinje analiza.
- ▶ Kako se analiza nastavlja, Local Run Manager zapisuje izlazne datoteke u mapu analize, a zatim te datoteke ponovno kopira u izlaznu mapu.












Mape Alignment (Usklađivanje)

Svaki put kad je ponovno na redu analiza, Local Run Manager stvara mapu Alignment (Usklađivanje) naziva **Alignment_N** (Usklađivanje_N), pri čemu je N neki broj u nizu.

Struktura mapa


 **Alignment (Usklađivanje)** – sadrži *.bam, *.vcf, FASTQ datoteke i datoteke specifične za modul za analizu.

 **Date and Time Stamp (Žig datuma i vremena)** – žig Datum_Vrijeme analize u obliku GGGGMMDD_HHMMSS



-  AnalysisError.txt
-  AnalysisLog.txt
-  aggregate.report.html
-  aggregate.report.pdf
-  aggregate.summary.csv
-  AmpliconCoverage_M#.tsv
-  AmpliconRunStatistics.xml
-  Sample1.genome.vcf.gz
-  Sample1.coverage.csv
-  Sample1.report.pdf
-  Sample1.summary.csv
-  Sample1.vcf.gz
-  Sample1.bam

 **FASTQ**

 **Sample1**

-  Sample1_L001_R1_001_fastq.gz

 **Stats (Statistički podaci)**

-  DemuxSummaryF1L1.txt
-  FastqSummaryF1L1.txt

 **Data**

 **Intensities**

 **BaseCalls**

 **L001** – sadrži *.bcl datoteke.

 **L001** – sadrži *.locs datoteke.

 **RTA Logs** – sadrži datoteke zapisnika analize softverom za RTA.

 **InterOp** – sadrži binarne datoteke korištene za izvješćivanje o metrici obrade sekvenciranja.

 **Logs (Zapisnici)** – sadrži datoteke zapisnika kojima se opisuju koraci izvedeni tijekom sekvenciranja.

RTAComplete.txt

RunInfo.xml

RunParameters.xml

Prepoznavanje baza i raznolikost indeksa

Kad se sekvenciraju uzorci na instrumentu NextSeq 550Dx, određivanjem baza definira se baza (A, C, G ili T) za svaki klaster određene pločice ili područje snimanja na protočnoj jedinici u određenom ciklusu. Instrument NextSeq 550Dx primjenjuje dvokanalno sekvenciranje koje zahtijeva samo dvije slike za šifriranje podataka za četiri baze DNA: jednu iz crvenog kanala i jednu iz zelenog kanala.

Postupak čitanja indeksa tijekom prepoznavanja baza razlikuje se od prepoznavanja baza tijekom drugih čitanja.

Čitanja indeksa moraju početi s bar jednom bazom koja nije G u nekom od prva dva ciklusa. Ako čitanje indeksa počinje s dva prepoznavanja baza G, ne generira se intenzitet signala. Signal mora biti prisutan u nekom od prva dva ciklusa da bi se osigurala radna svojstva demultipleksiranja.

Prilikom odabira indeksa tijekom stvaranja obrade pojavit će se upozorenje na nisku raznolikost ako indeksi ne zadovoljavaju preduvjete raznolikosti. Da biste spriječili pojavu upozorenja na nisku raznolikost, odaberite sekvence indeksa koje daju signal u oba kanala za svaki ciklus.

- ▶ Crveni kanal – A ili C
- ▶ Zeleni kanal – A ili T

Takav postupak prepoznavanja baza osigurava točnost prilikom analize niskopleksnih uzoraka. Da biste saznali više o sekvenciranju svojih indeksa, pogledajte informativni pregled za *TruSeq Custom Amplicon Kit Dx* (broj dokumenta: 1000000029772).

Tijekom stvaranja obrade u softveru Local Run Manager odabrat ćete broj uzoraka koji će se testirati. Softver automatski popunjava predložene kombinacije indeksa koje zadovoljavaju preduvjete raznolikosti indeksa. Iako nije obavezno upotrebljavati predložene kombinacije indeksa, to se preporučuje.

Povijest revizija

Dokument	Datum	Opis promjene
Broj dokumenta 1000000030329 v04	Kolovoz 2021.	Ažurirana adresa ovlaštenog predstavnika za EU.
Broj dokumenta 1000000030329 v03	Travanj 2020.	Ažurirana adresa ovlaštenog predstavnika za EU. Ažurirana adresa australskog sponzora.
Broj dokumenta 1000000030329 v02	Siječanj 2019.	Dodane informacije o v2.5 kompleta reagensa.
Broj dokumenta 1000000030329 v01	Kolovoz 2018.	Ažurirane regulatorne oznake.
Broj dokumenta 1000000030329 v00	Studeni 2017.	Početno izdanje.

Tehnička pomoć

Ako vam je potrebna tehnička pomoć, obratite se službi za tehničku podršku tvrtke Illumina.

Web-mjesto: www.illumina.com
Adresa e-pošte: techsupport@illumina.com

Telefonski brojevi službe za podršku korisnicima tvrtke Illumina

Regija	Besplatni telefon	Regionalno
Sjeverna Amerika	+1.800.809.4566	
Australija	+1.800.775.688	
Austrija	+43 800006249	+43 19286540
Belgija	+32 80077160	+32 34002973
Danska	+45 80820183	+45 89871156
Finska	+358 800918363	+358 974790110
Francuska	+33 805102193	+33 170770446
Hong Kong	800960230	
Irska	+353 1800936608	+353 016950506
Italija	+39 800985513	+39 236003759
Japan	0800.111.5011	
Kina	400.066.5835	
Nizozemska	+31 8000222493	+31 207132960
Njemačka	+49 8001014940	+49 8938035677
Norveška	+47 800 16836	+47 21939693
Novi Zeland	0800.451.650	
Singapur	+1.800.579.2745	
Španjolska	+34 911899417	+34 800300143
Švedska	+46 850619671	+46 200883979
Švicarska	+41 565800000	+41 800200442
Tajvan	00806651752	
Velika Britanija	+44 8000126019	+44 2073057197
Ostale države	+44.1799.534000	

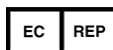
Sigurnosno-tehnički listovi – dostupni su na web-mjestu tvrtke Illumina na adresi support.illumina.com/sds.html.

Dokumentacija o proizvodima – dostupna je za preuzimanje u PDF obliku s web-mjesta tvrtke Illumina. Na adresi support.illumina.com odaberite proizvod, a zatim **Documentation & Literature** (Dokumentacija i literatura).



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 sponsor

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

ZA IN VITRO DIJAGNOSTIKU

© 2021. Illumina, Inc. Sva prava pridržana.

illumina®