

Local Run Manager Modul za analizo Germline Variant

Navodila za potek dela za NextSeq 550Dx

SAMO ZA IN VITRO DIAGNOSTIČNO UPORABO

Pregled	3
Vnos informacij o izvedbi	3
Metode analize	6
Ogled podatkov o izvedbi in vzorcu	7
Poročilo o analizi	7
Datoteke z rezultati analize	8
Zgodovina revizij	16
Tehnična pomoč	17



Ta dokument in vsebina v njem sta last družbe Illumina, Inc. in njenih podružnic (»Illumina«) ter sta namenjena le pogodbeno določeni uporabi njenih strank v povezavi z uporabo izdelkov, ki so opisani v tem dokumentu, in za noben drug namen. Tega dokumenta in vsebine v njem ne smete uporabljati ali distribuirati za kateri koli drug namen in/ali ju kakor koli drugače posredovati, razkriti ali razmnoževati brez predhodnega pisnega soglasja družbe Illumina. Illumina vam s tem dokumentom ne podeljuje nobene licence v okviru svojega patenta, blagovne znamke, avtorskih pravic ali pravic iz običajnega prava in nobenih podobnih pravic tretjih oseb.

Ustrezno kvalificirano in usposobljeno osebje mora natančno in dosledno upoštevati navodila v tem dokumentu, da zagotovi pravilno in varno uporabo izdelkov, opisanih v njem. Pred uporabo teh izdelkov morate v celoti prebrati vsebino tega dokumenta in se seznaniti z njo.

ČE NE PREBERETE VSEH NAVODIL V TEM DOKUMENTU IN JIH NE UPOŠTEVATE DOSLEDNO, LAHKO POVZROČITE OKVARO IZDELKOV, TELESNE POŠKODBE OSEB, VKLJUČNO Z UPORABNIKI IN DRUGIMI OSEBAMI, TER POŠKODBE DRUGE LASTNINE IN RAZVELJAVITE KAKRŠNO KOLI JAMSTVO, KI VELJA ZA IZDELKE.

ILLUMINA NE PREVZEMA NOBENE ODGOVORNOSTI ZA NEPRAVILNO UPORABO IZDELKOV, OPISANIH V TEM DOKUMENTU (VKLJUČNO Z NJIHOVIMI DELI IN PROGRAMSKO OPREMO).

© 2021 Illumina, Inc. Vse pravice so pridržane.

Vse blagovne znamke so last družbe Illumina, Inc. ali njihovih ustreznih lastnikov. Informacije o določenih blagovnih znamkah najdete na spletnem mestu www.illumina.com/company/legal.html.

Pregled

Modul Germline Variant Local Run Manager je namenjen uporabi s testom Illumina TruSeq Custom Amplicon Kit Dx. in instrumentom NextSeq 550Dx. Pri uporabi z modulom Germline Variant je test namenjen pripravi knjižnic, ki se uporabljajo za sekvenciranje DNK iz vzorcev periferne polne krvi.

Modul za analizo ocenjuje kratke regije pomnožene DNK, ali ampliconov, za variante. Usmerjeno sekvenciranje ampliconov omogoča visoko pokritost določenih regij. Glejte navodila za uporabo *TruSeq Custom Amplicon Kit Dx* (dokument št. 1000000029772).

Modul za analizo Germline Variant zahteva potrošni material za sekvenciranje za 300 ciklov. Za več informacij glejte navodila za uporabo kompleta reagentov *NextSeq 550Dx High Output Reagent Kit v2* ali *NextSeq 550Dx High Output Reagent Kit v2.5*.

O teh navodilih

Ta navodila vsebujejo navodila za nastavitve parametrov izvedbe za sekvenciranje in analizo za modul za analizo Germline Variant. Za informacije o nadzorni plošči Local Run Manager in nastavitvah sistema glejte *Referenčna navodila za NextSeq 550Dx instrument* (dokument št. 1000000009513).

Ogled Local Run Manager

Vmesnik Local Run Manager je prikazan v Operacijska programska oprema (NOS) NextSeq 550Dx ali prek spletnega brskalnika. Podprti spletni brskalnik je Chromium.



OPOMBA

Če uporabljate nepodprt brskalnik, prenesite podprti brskalnik, ko prejmete poziv v sporočilu "Confirm Unsupported Browser" (Potrdi nepodprt brskalnik). Izberite »here« (tukaj), da prenesete podprto različico brskalnika Chromium.

Ogled na monitorju instrumenta

- 1 Če si želite ogledati vmesnik Local Run Manager na monitorju instrumenta, izberite eno od naslednjih možnosti:
 - ▶ Na začetnem zaslonu NOS izberite **Local Run Manager**.
Kliknite X v zgornjem desnem kotu, da se vrnete v NOS, ko končate.
 - ▶ Izberite ikono NOS Minimize (Minimiraj), odprite spletni brskalnik Chromium, ki je na instrumentu, in v naslovno vrstico vnesite **http://localhost**.
Samo skrbniški uporabniki lahko minimizirajo NOS.

Ogled na omrežnem računalniku

- 1 Odprite spletni brskalnik Chromium v računalniku z dostopom do istega omrežja kot instrument in se povežite z uporabo naslova IP instrumenta ali imena instrumenta. Na primer, **http://myinstrument**.

Vnos informacij o izvedbi

Nastavitve parametrov

- 1 Prijavite se v Local Run Manager.

- 2 Izberite **Create Run** (Ustvari izvedbo) in izberite **Variant Germline** (Varianta zarodne linije).
- 3 Vnesite ime izvedbe, ki označuje izvedbo od sekvenciranja do končane analize.
Uporabite alfanumerične znake, presledke, podčrtaje ali pomišljaje.
- 4 **[Izbirno]** Vnesite opis izvedbe kot pomoč za opredelitev izvedbe.
Uporabite alfanumerične znake, presledke, podčrtaje ali pomišljaje.
- 5 Na spustnem seznamu izberite število vzorcev in nabor indeksov.
Pri izbiri upoštevajte naslednje informacije.
 - ▶ Na spustnem seznamu je število vzorcev z naborom indeksov. Na primer, 24-Set 1 označuje 24 vzorcev, ki jih je treba testirati, z indeksi iz indeksnega nabora 1.
 - ▶ Številke naborov indeksov se nanašajo na različne nize indeksov i5. Nabor 1 in nabor 2 zagotavljata raznolikost indeksov. Na voljo sta dva nabora indeksov, ki pomagata preprečiti deplecijo enega nabora.
 - ▶ Izberite število vzorcev, ki je najbližje številu vzorcev, ki jih testirate. Če natančnega števila vzorcev ni na seznamu, izberite najbližjo številko, vendar manj od števila, ki jih testirate. Če želite na primer testirati 18 vzorcev, izberite 16 vzorcev.
 - ▶ Kombinacije vdolbinic za Sample (Vzorec) in indeksov, ki izpolnjujejo zahteve za raznolikost indeksov, so označene z zeleno barvo. Če izberete druge kombinacije vdolbinic in indeksov, ko shranite izvedbo, prejmete opozorilo, če zahteve za raznolikost indeksov niso izpolnjene.

Uvoz manifestnih datotek za izvedbo

- 1 Prepričajte se, da so manifesti, ki jih želite uvoziti, na voljo na dostopnem omrežnem mestu ali na pogonu USB.
- 2 Izberite **Import Manifests** (Uvozi manifeste).
- 3 Pomaknite se do manifestne datoteke in izberite manifeste, ki jih želite dodati.



OPOMBA

Če želite, da so manifestne datoteke na voljo za vse izvedbe z modulom za analizo Germline Variant, dodajte manifeste s funkcijo Module Settings (Nastavitve modula). Ta funkcija zahteva dovoljenja za raven skrbnika. Za več informacij glejte *Referenčna navodila za NextSeq 550Dx instrument (dokument št. 1000000009513)*.

Določitev vzorcev za izvedbo

Vzorce za izvedbo določite z eno od možnosti in navodili, ki sledijo.

- ▶ **Ročni vnos vzorcev** - na zaslonu Create Run (Ustvari izvedbo) uporabite prazno tabelo.
- ▶ **Uvoz vzorcev** - pomaknite se v zunanjo datoteko v formatu z vrednostmi, ločenimi z vejicami (*.csv). Predloga je na voljo za prenos na zaslonu Create Run (Ustvari izvedbo).

Ko izpolnite tabelo vzorcev, lahko izvozite podatke o vzorcih v zunanjo datoteko. Uporabite datoteko kot referenco, ko pripravljate knjižnice ali uvažate datoteko za drugo izvedbo.

Ročni vnos vzorcev

- 1 Vnesite enolični ID vzorca v polje Sample ID (ID vzorca).
Uporabite alfanumerične znake, pomišljaje ali podčrtaje.

- 2 **[Izbirno]** Za pozitivne ali negativne kontrolne vzorce kliknite desno tipko miške in izberite vrsto kontrole.
- 3 **[Izbirno]** V polje Sample Description (Opis vzorca) vnesite opis vzorca.
Uporabite alfanumerične znake, pomišljaje ali podčrtaje.
Opisi vzorcev so povezani z imenom vzorca. Opisi vzorcev se prepišejo, če se isto ime vzorca znova uporabi v poznejši izvedbi.
- 4 Izberite adapter indeksa 1 s spustnega seznama Index 1 (i7).
Ko uporabljate predlagane vzorčne vdolbinice, programska oprema samodejno napolni adapterje indeksa i7 in i5, ki izpolnjujejo zahteve za raznolikost indeksa. Če na seznamu ni natančnega števila vzorcev, ki jih testirate, izberite indeksne adapterje za dodatne vdolbinice. Če morate izbrati indekse za dodatne vdolbinice ali če ne uporabljate priporočenih kombinacij indeksnih adapterjev, pred izbiro indeksov obvezno preberite razdelek *Določitev baze in raznolikost indeksov na strani 14*.
- 5 Izberite Index 2 adapter (Adapter indeksa 2) s spustnega seznama Index 2 (i5) (Indeks 2 (i5)).
- 6 Izberite manifestno datoteko s spustnega seznama Manifest.
- 7 Izberite možnost, da si ogledate, natisnete ali shranite postavitev ploščice kot referenco za pripravo knjižnic:
 - ▶ Izberite ikono  **Print** (Natisni), da se prikaže postavitev ploščice. Izberite **Print** (Natisni) za tiskanje postavitve ploščice.
 - ▶ Izberite **Export** (Izvoz) za izvoz podatkov o vzorcu v zunanjo datoteko.Prepričajte se, da so informacije o manifestu in vzorcu pravilne. Nepravilne informacije lahko vplivajo na rezultate.
- 8 Izberite **Save Run** (Shrani izvedbo).

Uvoz vzorcev

- 1 Izberite **Import Samples** (Uvoz vzorcev) in poiščite mesto datoteke CSV. Obstajata dve vrsti datotek, ki jih lahko uvozite.
 - ▶ Izberite **Template** (Predloga) na zaslonu Create Run (Ustvari izvedbo), da ustvarite novo postavitev ploščice. Datoteka predloge vsebuje pravilne naslove stolpcev za uvoz. Vnesite informacije o vzorcu v vsak stolpec za vzorce v izvedbi. Izbrišite podatke v neuporabljenih celicah in nato shranite datoteko.
 - ▶ Uporabite datoteko z informacijami o vzorcih, ki ste jo izvozili iz modula Germline Variant s funkcijo Export (Izvoz).
- 2 Izberite ikono  **Print** (Natisni), da se prikaže postavitev ploščice.
- 3 Izberite **Print** (Natisni), da natisnete postavitev ploščice kot referenco za pripravo knjižnic.
- 4 **[Izbirno]** Izberite **Export** (Izvoz) za izvoz informacij o vzorcu v zunanjo datoteko.
Prepričajte se, da so informacije o manifestu in vzorcu pravilne. Nepravilne informacije lahko vplivajo na rezultate.
- 5 Izberite **Save Run** (Shrani izvedbo).

Urejanje izvedbe

Za navodila o urejanju informacij v vaši izvedbi pred sekvenciranjem glejte *Referenčna navodila za NextSeq 550Dx instrument (dokument št. 1000000009513)*.

Metode analize

Modul za analizo Germline Variant izvede naslednje korake analize in nato zapiše izhodne datoteke analize v mapo Alignment.

- ▶ Demultipleksira indeksne odčitke
- ▶ Ustvari datoteke FASTQ
- ▶ Poravna z referenco
- ▶ Identificira variante

Demultipleksiranje

Demultipleksiranje primerja vsako sekvenco indeksnega odčitavanja z indeksnimi sekvencami, določenimi za izvedbo. V tem koraku se ne upoštevajo nobene vrednosti kakovosti.

Indeksni odčitki so označeni po naslednjih korakih:

- ▶ Vzorci so oštevilčeni z začetkom od 1 na podlagi vrstnega reda, v katerem so navedeni za izvedbo.
- ▶ Številka vzorca 0 je rezervirana za gruče, ki niso bile dodeljene vzorcu.
- ▶ Gruče so dodeljene vzorcu, ko se sekvenca indeksov točno ujema ali ko je na voljo največ eno neujemanje na indeksno odčitavanje.

Ustvarjanje datoteke FASTQ

Po demultipleksiranju programska oprema ustvari datoteke vmesne analize v obliki zapisa FASTQ, ki je besedilna oblika, uporabljena za predstavitev zaporedij. Datoteke FASTQ vsebujejo odčitke za vsak vzorec in povezane ocene kakovosti. Gruče, ki niso prešle skozi filter, so izključene.

Vsaka datoteka FASTQ vsebuje odčitke samo za en vzorec, ime tega vzorca pa je vključeno v ime datoteke FASTQ. Datoteke FASTQ so primarni vhod za poravnavo. Na vzorec se ustvari osem datotek FASTQ, štiri iz odčitka 1 in štiri iz odčitka 2.

Poravnava

Med korakom poravnave vezani algoritem Smith-Waterman poravna gruče iz vsakega vzorca s sekvencami amplikona, določenimi v manifestni datoteki.

Vezani algoritem Smith-Waterman izvaja polglobalne poravnave sekvenc, da določi podobne regije med dvema sekvencama. Namesto primerjanja celotne sekvence algoritem Smith-Waterman primerja segmente vseh mogočih dolžin.

Vsako odčitavanje z obeh koncev se oceni glede na njegovo poravnavo z ustreznimi sekvencami sond za to odčitavanje.

- ▶ Odčitek 1 se oceni glede na protismerni komplement navzdolnjih za lokus specifičnih oligonukleotidov (Downstream Locus-Specific Oligos - DLSO).
- ▶ Odčitek 2 se oceni glede na navzgorne za lokus specifične oligonukleotide (Upstream Locus-Specific Oligos - ULSO).
- ▶ Če se začetek odčitka ujema s sekvenco sonde z največ tremi razlikami (neujemanja ali zamiki zaradi vodilnih indelov), je celotna dolžina odčitka poravnana s tarčo amplikona za to sekvenco.
- ▶ Indelov znotraj DLSO in ULSO zaradi kemije testa ne opazimo.

Poravnave se filtrirajo iz rezultatov poravnave na podlagi deležev neujemanja v regiji zanimanja ali polnem ampliconu, odvisno od dolžine amplicona. Filtrirane poravnave so zapisane v datotekah poravnave kot neporavnane in se ne uporabljajo v določanju variant.

Določanje variant

Orodje za določanje variant Pisces, ki ga je razvila družba Illumina, identificira variante, ki so prisotne v vzorcu DNK.

Orodje za določanje variant Pisces identificira SNV-je, MNV-je in majhne indele v 3 korakih:

- ▶ vsak položaj v referenčnem genomu obravnava ločeno;
- ▶ šteje baze na danem položaju za poravnane odčitke, ki prekrivajo položaj;
- ▶ izračuna oceno variante, ki meri kakovost določitve z uporabo Poissonovega modela. Izključene so variante z oceno kakovosti pod Q20.

Če varianta preide vse filtre, je označena kot PASS (PREHOD) v VCF.

Za več informacij glejte github.com/Illumina/Pisces/wiki.

Ogled podatkov o izvedbi in vzorcu

- 1 Na nadzorni plošči Local Run Manager kliknite ime izvedbe.
- 2 Na zavihku Run Overview (Pregled izvedbe) preglejte metrike izvedbe sekvenciranja.
- 3 **[Izbirno]** Za kopiranje poti do izhodne mape izvedbe kliknite ikono  **Copy to Clipboard** (Kopiraj v odložišče).
- 4 Za pregled parametrov izvedbe in informacij o potrošnem materialu kliknite zavihek Sequencing Information (Informacije o sekvenciranju).
- 5 Z ogled lokacije poročila o analizi kliknite zavihek Samples and Results (Vzorci in rezultati).
 - ▶ Če je bila analiza ponovljena, razširite spustni seznam Select Analysis (Izberi analizo) in izberite ustrezno analizo.
- 6 Za kopiranje poti do mape Analysis kliknite ikono  **Copy to Clipboard** (Kopiraj v odložišče).

Če želite več informacij o zavihkih Run Overview (Pregled izvedbe) in Sequencing Information (Informacije o sekvenciranju) in kako znova uvrstiti analizo v čakalno vrsto, glejte *Referenčna navodila za NextSeq 550Dx instrument (dokument št. 1000000009513)*.

Poročilo o analizi

Rezultati analize so povzeti v zavihku Samples and Results (Vzorci in rezultati) in kot zbirno poročilo v mapi Alignment. Poročilo za vsak vzorec je na voljo tudi v obliki zapisa datoteke PDF za vsak vzorec.

Informacije o zavihkih Vzorci in Rezultati

- 1 Kliknite vzorec na seznamu, da si ogledate poročilo o vzorcu.

Preglednica 1 Informacije o izvedbi in vzorcu

Naslov stolpca	Opis
Run Status (Stanje izvedbe)	Navaja, ali je bilo sekvenciranje uspešno ali neuspešno.
Total Yield (GB) (Skupni donos (GB))	Število določenih baz v sekvenciranju. Prikazuje prag za prehod in stanje uspešnosti ali neuspešnosti.
% ≥ Q30	Odstotek odčitkov pri izvedbi sekvenciranja z oceno kakovosti 30 (Q30) ali več. Prikazuje prag za prehod in stanje uspešnosti ali neuspešnosti.
Sample ID (ID vzorca)	ID vzorca, ki je bilo navedeno ob ustvarjanju izvedbe.
Total PF Reads (Skupno število odčitkov PF)	Skupno število odčitkov, ki so prešli skozi filter.
Read 1% ≥ Q30 (Odčitavanje 1 % ≥ Q30)	Odstotek odčitkov v odčitavanju 1 z oceno kakovosti 30 (Q30) ali več za vzorec.
Read 2% ≥ Q30 (Odčitavanje 2 % ≥ Q30)	Odstotek odčitkov v odčitavanju 2 z oceno kakovosti 30 (Q30) ali več za vzorec.
Autosome call rate (Delež avtosomnih določitev)	Število genomskega položaja v avtosomih (kromosomi od 1 do 22), ki izpolnjujejo vnaprej določen prag vrednosti zaupanja, deljeno s skupnim številom preiskanih avtosomnih genomskega položaja. Stopnja določitev je opisana na podlagi posameznega vzorca in se poroča kot odstotek, ki se izračuna kot 1 minus (število avtosomnih položajev z nepopolnimi določitvami, deljeno s skupnim številom sekvenciranih avtosomnih položajev).

Preglednica 2 Informacije v Poročilu o Vzorcih

Naslov stolpca	Opis
Vzorec	ID vzorca, ki je bilo navedeno ob ustvarjanju izvedbe.
Report Date (Datum poročila)	Datum nastanka poročila.
Sample Information (Informacije o vzorcu)	ID vzorca, ki je bil podan ob ustvarjanju izvedbe, skupno št. odčitkov, ki so prešli skozi filter v vzorcu, odstotek odčitkov za vzorec z oceno kakovosti 30 (Q30) ali več in delež avtosomnih določitev.
Amplicon Summary (Povzetek o amplikonu)	Skupno število sekvenciranih regij amplikona in skupna dolžina v baznih parih sekvenciranih amplikonov v tarčnih regijah, za vzorec in manifestno datoteko. Manifestna datoteka določa referenčni genom in tarčna referenčna območja, ki se uporabljajo v koraku poravnave.
Read Level Statistics (Statistika ravni določitev)	Število in odstotek odčitkov za vzorec, ki pokriva vsak položaj v referenci, za odčitavanje 1 in odčitavanje 2.
Variants Summary (Povzetek variant)	Število SNV-jev, insercij in delecij, zaznanih za vzorec, ki so prešle predlagane vrednosti, da se ugotovi, ali so rezultati kakovosti znotraj sprejemljivega razpona.
Coverage Summary (Povzetek pokritosti)	Skupno število poravnanih baz, deljeno z velikostjo tarčne regije, in odstotek amplikonskih regij z vrednostmi pokritosti, ki so večje od spodnjega praga pokritosti $0,2 \times$ srednje pokritosti amplikona, za vzorec.
Coverage Plots (Graf pokritosti)	Graf pokritosti po regijah amplikonov prikazuje pokritost za vse regije amplikonov za vzorec. Regije z vrednostmi pokritosti, nižjimi od praga pokritosti, so označene z rdečo. Povprečje vseh vrednosti je označeno z oranžno črto.
Software Versions (Različice programske opreme)	Različice programske opreme, ko je bil vzorec sekvenciran. Vključuje Operacijska programska oprema (NOS) NextSeq 550Dx, programsko opremo Local Run Manager, programsko opremo RTA in različico modula Germline Variant.

Datoteke z rezultati analize

Za modul za analizo Germline Variant se ustvarijo naslednje izhodne datoteke analize in zagotavljajo rezultate analize za poravnavo in določanje variant. Datoteke z rezultati analize so v mapi Alignment.

Ime datoteke	Opis
Demultipleksiranje (*.txt)	Vmesne datoteke, ki vsebujejo rezultate povzetka demultipleksiranja.
FASTQ (*.fastq.gz)	Vmesne datoteke, ki vsebujejo kakovostne določitve baze. Datoteke FASTQ so primarni vhod za korak poravnave.
Datoteke poravnave v obliki zapisa BAM (*.bam)	Vsebuje poravnane odčitke za dani vzorec.
Datoteke določanja variant v obliki zapisa genom VCF (*.genome.vcf.gz)	Vsebuje genotip za vsak položaj, bodisi določen kot varianta bodisi določen kot referenca.
Datoteke določitve variant v obliki zapisa VCF (*.vcf.gz)	Vsebuje vse variante, določene v tarčni regiji.
AmpliconCoverage_M1.tsv	Vsebuje informacije o pokritosti na ampikon na vzorec za vsak predloženi manifest. M# predstavlja številko manifesta.

Oblika zapisa datoteke demultipleksiranja

Proces demultipleksiranja odčita indeksno sekvenco, ki se drži vsake gruče, da določi, iz katerega vzorca gruča izvira. Kartiranje med gručami in številka vzorca so zapisani v datoteko za demultipleksiranje (*.demux) za vsako ploščico pretočne celice.

Oblika zapisa poimenovanja datoteke demultipleksiranja je **s_1_X.demux**, pri čemer je X številka ploščice.

Datoteka demultipleksiranja se začne z glavo:

- ▶ Različica (4-bajtno celo število), trenutno 1
- ▶ Število gruč (4-bajtno celo število)

Preostanek datoteke je sestavljen iz številke vzorcev za vsako gručo s ploščice.

Ko je korak demultipleksiranja končan, programska oprema ustvari datoteko demultipleksiranja z imenom **DemultiplexSummaryF1L1.txt**.

- ▶ V imenu datoteke **F1** predstavlja številko pretočne celice.
- ▶ V imenu datoteke **L1** predstavlja številko pasu.
- ▶ Rezultat demultipleksiranja je tabela z 1 vrstico na ploščico in 1 stolpcem na vzorec, vključno z vzorcem 0.
- ▶ Najpogostejše sekvence v indeksnih odčitkih.

Oblika zapisa datoteke FASTQ

FASTQ je besedilna oblika datoteke, ki vsebuje določitve baz in vrednosti kakovosti na odčitek. Vsak zapis vsebuje 4 vrstice:

- ▶ Identifikator
- ▶ Sekvenca
- ▶ Znak plus (+)
- ▶ Oceno kakovosti Phred v kodiranem formatu ASCII + 33

Identifikator je oblikovan kot:

@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber

Primer:

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

Oblika zapisa datoteke BAM

Datoteka BAM (*.bam) je stisnjena dvojiška različica datoteke SAM, ki se uporablja za predstavitev usklajenih sekvenc do 128 Mb. Obliki zapisa SAM in BAM sta podrobno opisani na spletni strani <https://samtools.github.io/hts-specs/SAMv1.pdf>.

Datoteke BAM uporabljajo obliko zapisa za poimenovanje datotek **SampleName_S#.bam**, pri čemer je # številka vzorca, določena z vrstnim redom, po katerem so vzorci navedeni za izvedbo.

Datoteke BAM vsebujejo razdelek z glavo in razdelek s poravnavo:

- ▶ **Glava** - vsebuje informacije o celotni datoteki, kot so ime vzorca, dolžina vzorca in način poravnave. Poravnave v odseku poravnave so povezane s posebnimi informacijami v odseku glave.
- ▶ **Poravnave** - vsebuje ime odčitavanja, sekvenco odčitavanja, kakovost odčitavanja, informacije o poravnavi in oznake po meri. Ime odčitka vključuje kromosom, začetno koordinato, kakovost poravnave in niz deskriptorja ujemanja.

Razdelek poravnave vsebuje naslednje informacije za vsak odčitek ali par odčitkov:

- ▶ **AS:** Kakovost obojestranske poravnave
- ▶ **RG:** Skupina odčitkov, ki označuje število odčitkov za določen vzorec.
- ▶ **BC:** Oznaka črtne kode, ki označuje ID demultipleksiranega vzorca, povezanega z odčitkom.
- ▶ **SM:** Kakovost enostranske poravnave.
- ▶ **XC:** Deskriptorski niz ujemanja.
- ▶ **XN:** Oznaka imena amplikona, ki beleži ID amplikona, povezanega z odčitkom

Datoteke z indeksom BAM (*.bam.bai) navajajo indeks ustrezne datoteke BAM.

Oblika zapisa datoteke VCF

Oblika zapisa za določanje variant (Variant Call Format - VCF) je pogosta oblika datoteke, ki jo je razvila znanstvena skupnost genomike. Vsebuje informacije o variantah, najdenih na določenih položajih v referenčnem genomu. Datoteke VCF se končajo s pripono .vcf

Glava datoteke VCF vključuje različico oblike datoteke VCF in različico orodja za določanje variant ter navaja anotacije, uporabljene v preostanku datoteke. Glava VCF vključuje tudi datoteko referenčnega genoma in datoteko BAM. Zadnja vrstica v naslovu vsebuje naslove stolpcev za podatkovne vrstice. Vsaka podatkovna vrstica datoteke VCF vsebuje informacije o eni varianti.

Glave datotek VCF

Glava	Opis
CHROM	Kromosom referenčnega genoma. Kromosomi so prikazani v enakem vrstnem redu kot referenčna datoteka FASTQ.
POS	Položaj ene baze variante v referenčnem kromosomu. Za SNP je ta položaj referenčna baza z varianto; za indele ali delecije je ta položaj referenčna baza tik pred varianto.
ID	Številka rs za varianto, pridobljena iz dbSNP.txt, če je ustrezno. Če na tej lokaciji obstaja več števil rs, je seznam razmejen s podpičji. Če na tem položaju obstaja vnos dbSNP, se uporabi znak za manjkajočo vrednost ('.').
REF	Referenčni genotip. Na primer, delecija enega T je predstavljena kot referenčni TT in nadomestni T. Varianta posameznega nukleotida A v T je predstavljena kot referenca A in nadomestna T.
ALT	Aleli, ki se razlikujejo od referenčnega odčitka. Na primer, insercija enega T je predstavljena kot referenčni A in nadomestni AT. Varianta posameznega nukleotida A v T je predstavljena kot referenčni A in nadomestni T.
QUAL	Ocena kakovosti, ki jo je dodelilo orodje za določanje variant in razširil algoritem Phred. Višji rezultati kažejo večje zaupanje v varianto in manjšo verjetnost napak. Za oceno kakovosti Q je ocenjena verjetnost napake $10^{-(Q/10)}$. Na primer, nabor identifikacij Q30 ima 0,1-odstotno stopnjo napak. Mnogo orodij za določanje variant dodeli ocene kakovosti na podlagi svojih statističnih modelov, ki so visoki glede na stopnjo ugotovljenih napak.

Anotacije datoteke VCF

Glava	Opis
FILTER	Če je prehod skozi vse filtre uspešen, je v stolpcu filtra zapisano PASS (PREHOD). LowDP - uporablja se za mesta z globino pokritosti pod 150 x. Za položaje ampikona, ki jih pokrivata tako smerni kot tudi protismerni odčitek, je to enakovredno 300 prekrivajočim odčitkom z obeh koncev. q20 - ocena kakovosti < 20. MultiAllelicSite - varianta ne ustreza diploidnemu modelu. R5x9 - število sosednjih ponovitev (dolžine od 1 do 5 bp) na določitve variant ≥ 9. SB - odstopanje verige je višje od danega praga.
INFO	Možni vnosi v stolpcu INFO vključujejo: AC - število alelov v genotipih za vsak alel ALT v enakem vrstnem redu, kot je navedeno. AF - pogostnost alelov za vsak alel ALT v enakem vrstnem redu, kot je navedeno. AN - skupno število alelov v določenih genotipih. CD - opozorilo, ki označuje, da se SNP pojavlja v regiji kodiranja vsaj 1 vnosa RefGene. DP - globina (število določitve baz, poravnanih s položajem in uporabljenih v določanju variant). Exon - seznam regij eksonov, ločenih z vejico, ki so bile odčitane iz vnosa RefGene. FC - funkcionalna posledica. GI - seznam genskih ID-jev, ločenih z vejico, ki so bili odčitani iz vnosa RefGene. QD - zaupanje v varianto/kakovost glede na globino. TI - seznam transkriptnih ID-jev, ločenih z vejico, ki so bili odčitani iz vnosa RefGene.
FORMAT	Stolpec formata navaja polja, ločena z dvopičji. Na primer GT:GQ. Seznam polj je odvisen od uporabljenega orodja za določanje variant. Razpoložljiva polja vključujejo: AD - vnos obrazca X, Y, pri čemer je X število izpisov reference, Y pa število alternativnih izpisov. DP - približna globina odčitavanja; odčitavanja z MQ = 255 ali z neustreznimi ujemanji se filtrirajo. GQ - kakovost genotipa. GT - genotip. 0 ustreza referenčni bazi, 1 ustreza prvemu vnosu v stolpcu ALT itd. Desna poševnica (/) pomeni, da ni na voljo nobenih informacij o faziranju. NC - delež baz, ki niso bile določene ali imajo kakovost določitve baze pod minimalnim pragom. NL - raven hrupa; ocena hrupa pri določanju baz v tem položaju. SB - odstopanje verig na tem položaju. Večje negativne vrednosti kažejo na manj odstopanja; vrednosti blizu 0 kažejo na več odstopanja. VF - frekvenca variante; odstotek odčitavanj, ki podpirajo alternativni alel.
SAMPLE	Stolpec z vzorcem vsebuje vrednosti, navedene v stolpcu FORMAT.

Datoteke VCF genoma

Datoteke VCF genoma (gVCF) so datoteke VCF v4.1, ki sledijo nizu konvencij za predstavljanje vseh mest v genomu v razumno kompaktnem formatu. Datoteke gVCF (*.genome.vcf.gz) vključujejo vsa mesta znotraj regije zanimanja v eni datoteki za vsak vzorec.

Datoteka gVCF prikazuje ne-določitve na položajih, ki ne preidejo skozi vse filtre. Oznaka genotipa (GT) ./, označuje ne-določitve.

Za več informacij glejte sites.google.com/site/gvcftools/home/about-gvcf.

Datoteka pokritosti ampikona

Za vsako manifestno datoteko se ustvari datoteka pokritosti ampikona. M# v imenu datoteke predstavlja številko manifesta.

Vsaka datoteka vključuje vrstico z glavo, ki vsebuje ID-je vzorca, povezane z manifestom. Vsaka datoteka vsebuje naslednje informacije.

- ▶ ID tarče, kot je naveden v manifestu.
- ▶ Globina pokritosti odčitkov, ki prehajajo skozi filter.

Dodatne izhodne datoteke

Naslednje izhodne datoteke zagotavljajo dodatne informacije ali povzemajo rezultate izvedbe in napake pri analizi. Čeprav te datoteke niso potrebne za ocenjevanje rezultatov analize, jih je mogoče uporabiti za odpravljanje težav. Vse datoteke so v mapi Alignment, razen če ni navedeno drugače.

Ime datoteke	Opis
AnalysisLog.txt	Dnevnik obdelave, v katerem je opisan vsak korak, do katerega je prišlo med analizo mape trenutne izvedbe. Ta datoteka ne vsebuje sporočil o napakah. Nahaja se v mapi Alignment.
AnalysisError.txt	Dnevnik obdelave, v katerem so navedene vse napake, do katerih je prišlo med analizo. Če ne pride do napak, je ta datoteka prazna. Nahaja se v mapi Alignment.
DemultiplexSummaryF1L1#.txt	Poročila o rezultatih demultipleksiranja v tabeli z 1 vrstico na ploščico in 1 stolpcem na vzorec. Številka predstavlja pas 1, 2, 3 ali 4 pretočne celice. Nahaja se v mapi Alignment.
AmpliconRunStatistika.xml	Vsebuje povzetek statističnih podatkov, specifičnih za izvedbo. Nahaja se v mapi Alignment.

Analysis Folder (Mapa analize)

V mapi analize so datoteke, ki jih ustvari programska oprema Local Run Manager.

Razmerje med izhodno mapo in mapo analize je povzeto takole:

- ▶ Med sekvenciranjem Real-Time Analysis (RTA) napolni izhodno mapo z datotekami, ustvarjenimi med analizo slike, določanjem baz in ocenjevanjem kakovosti.
- ▶ RTA kopira datoteke v mapo analize v realnem času. Ko RTA dodeli oceno kakovosti vsaki bazi za vsak cikel, programska oprema zapiše datoteko RTAComplete.txt v obe mapi.
- ▶ Ko je prisotna datoteka RTAComplete.txt, se začne analiza.
- ▶ Ko se analiza nadaljuje, Local Run Manager zapisuje izhodne datoteke v mapo analize in nato kopira datoteke nazaj v izhodno mapo.

Mape poravnave

Ob vsaki ponovni uvrstitvi analize v čakalno vrsto Local Run Manager ustvari mapo poravnave z imenom **Alignment_N**, pri čemer je N zaporedna številka.

Struktura mape

 **Alignment** - vsebuje datoteke *.bam, *.vcf, FASTQ in datoteke, specifične za modul za analizo.

 **Date and Time Stamp** - datum_čas žig analize kot LLLLMMDD_HHMMSS

 AnalysisError.txt

 AnalysisLog.txt

 aggregate.report.html

- aggregate.report.pdf
 - aggregate.summary.csv
 - AmpliconCoverage_M#.tsv
 - AmpliconRunStatistics.xml
 - Sample1.genome.vcf.gz
 - Sample1.coverage.csv
 - Sample1.report.pdf
 - Sample1.summary.csv
 - Sample1.vcf.gz
 - Sample1.bam
 - FASTQ
 - Sample1
 - Sample1_L001_R1_001_fastq.gz
 - Stats
 - DemuxSummaryF1L1.txt
 - FastqSummaryF1L1.txt
- Data
 - Intensities
 - BaseCalls
 - L001 - vsebuje datoteke *.bcl.
 - L001 - vsebuje datoteke *.locs.
 - RTA Logs - vsebuje dnevniške datoteke iz analize programske opreme RTA.
- InterOp - vsebuje binarne datoteke, ki se uporabljajo za poročanje o metrikah sekvenciranja.
- Logs - vsebuje dnevniške datoteke, ki opisujejo korake, izvedene med sekvenciranjem.
- RTAComplete.txt
- RunInfo.xml
- RunParameters.xml

Določitev baze in raznolikost indeksov

Ko so vzorci sekvencirani na instrumentu NextSeq 550Dx, se z določitvijo baze določi baza (A, C, G ali T) za vsako gručo določene ploščice ali območja slikanja na pretočni celici v določenem ciklu. Instrument NextSeq 550Dx uporablja 2-kanalno sekvenciranje, kjer sta za kodiranje podatkov za štiri baze DNK potrebni le dve sliki - 1 iz rdečega kanala in ena iz zelenega kanala.

Postopek indeksnih odčitavanj za določanje baz se razlikuje od določanja baz med drugimi odčitavanji.

Indeksna odčitavanja se morajo začeti z vsaj eno bazo, ki ni G, v katerem koli od prvih dveh ciklov. Če se indeksno odčitavanje začne z dvema določitvama baze G, se intenziteta signala ne ustvari. Če želite zagotoviti učinkovitost demultipleksiranja, mora biti prisoten signal v katerem koli od prvih dveh ciklov.

Če indeksi ne izpolnjujejo zahtev glede raznolikosti, se pri izbiranju indeksov med ustvarjanjem izvedbe prikaže opozorilo o nizki raznolikosti. Da bi preprečili opozorilo o nizki raznolikosti, izberite sekvence indeksov, ki zagotavljajo signal v obeh kanalih za vsak posamezen cikel.

- Rdeči kanal - A ali C
- Zeleni kanal - A ali T

Ta postopek določanja baz zagotavlja točnost pri analiziranju vzorcev z nizko stopnjo multipleksiranja (low-plex). Če želite več informacij o sekvencah indeksov, glejte navodila za uporabo *TruSeq Custom Amplicon Kit Dx* (dokument št. 1000000029772).

Med ustvarjanjem v programski opremi Local Run Manager boste izbrali število vzorcev, ki jih boste testirali. Predlagane kombinacije indeksov, ki izpolnjujejo zahteve glede raznolikosti indeksov, samodejno izpolni programska oprema. Čeprav vam ni treba uporabiti predlaganih kombinacij indeksov, je to priporočljivo.

Zgodovina revizij

Dokument	Datum	Opis spremembe
Dokument št. 1000000030329 v04	Avgust 2021	Posodobljen naslov pooblaščenega predstavnika v EU.
Dokument št. 1000000030329 v03	December 2019	Posodobljen naslov pooblaščenega predstavnika v EU. Posodobljen naslov avstralskega sponzorja.
Dokument št. 1000000030329 v02	Januar 2019	Dodane informacije o kompletih reagentov v2.5.
Dokument št. 1000000030329 v01	Avgust 2018	Posodobljene regulativne oznake.
Dokument št. 1000000030329 v00	November 2017	Prva izdaja.

Tehnična pomoč

Za tehnično pomoč se obrnite na tehnično podporo družbe Illumina.

Spletno mesto: www.illumina.com
E-pošta: techsupport@illumina.com

Telefonske številke tehnične podpore družbe Illumina

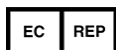
Regija	Brezplačna št.	Regionalna št.
Severna Amerika	+1 800 809 4566	
Avstralija	+1 800 775 688	
Avstrija	+43 800006249	+43 19286540
Belgija	+32 80077160	+32 34002973
Kitajska	400 066 5835	
Danska	+45 80820183	+45 89871156
Finska	+358 800918363	+358 974790110
Francija	+33 805102193	+33 170770446
Nemčija	+49 8001014940	+49 8938035677
Hongkong	800960230	
Irska	+353 1800936608	+353 016950506
Italija	+39 800985513	+39 236003759
Japonska	0800.111.5011	
Nizozemska	+31 800 022 2493	+31 207 132 960
Nova Zelandija	0800 451 650	
Norveška	+47 800 16836	+47 21939693
Singapur	+1 800 579 2745	
Španija	+34 911899417	+34 800300143
Švedska	+46 850619671	+46 200883979
Švica	+41 565800000	+41 800200442
Tajvan	00806651752	
Združeno kraljestvo	+44 8000126019	+44 2073057197
Druge države	+44 1799 534000	

Varnostni listi - na voljo na spletnem mestu družbe Illumina na naslovu support.illumina.com/sds.html.

Dokumentacija izdelka - na voljo za prenos v obliki zapisa PDF na spletnem mestu družbe Illumina. Pojdite na support.illumina.com, izberite izdelek, nato pa izberite razdelek **Dokumentacija in literatura**.



Illumina
5200 Illumina Way
San Diego, California 92122
ZDA
+1.800.809.ILMN (4566)
+1.858.202.4566
(zunaj Severne Amerike)
techsupport@illumina.com
www.illumina.com



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

Sponzor za Avstralijo

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

SAMO ZA IN VITRO DIAGNOSTIČNO UPORABO

© 2021 Illumina, Inc. Vse pravice so pridržane.

illumina®