

DRAGEN for Illumina DNA Prep with Enrichment Dx

NovaSeq 6000Dx:n tuotedokumentaatio

Versiohistoria

Asiakirja	Päivämäärä	Muutoksen kuvaus
200014776 v02	Syyskuu 2022	Korjattu manifestitiedoston muoto tekstitiedostosta (*.txt) BED-tiedostoon (*.bed) ajojen luomishjeissa. Korjattu VCF-konsensustiedostot VFC-tiedostoiksi analyysin tuotosta käsittelevässä osiossa.
200014776 v01	Elokuu 2022	Lisätty: Asetukset-osio. Systemaattisen kohinan suodatus-osio. Päivitetty ajon luomisen ohjeita sisältämään enemmän yksityiskohtia. Korjattu kirjoitusvirhe- ja kielioppivirheitä. Täsmennetty, että ohjeet on tarkoitettu sovellusta varten, kun sitä käytetään NovaSeq 6000Dx Instrument -laitteen kanssa. Päivitetty tiedot VCF-tulostetiedoston sisällöstä.
200014776 v00	Maaliskuu 2022	Ensimmäinen versio.

Tämä asiakirja ja sen sisältö ovat Illumina, Inc:n ja sen tytäryhtiöiden ("Illumina") omaisuutta, ja ne on tarkoitettu ainoastaan Illuminan asiakkaiden sopimuskäyttöön tässä kuvattujen tuotteiden käyttöön liittyen eikä mihinkään muuhun tarkoitukseen. Tätä asiakirjaa ja sen sisältöä ei saa käyttää tai jakaa missään muussa tarkoituksessa ja/tai välittää, paljastaa tai jäljentää millään muulla tavoin ilman Illuminalta ennakoon saatua kirjallista lupaa. Illumina ei tällä asiakirjalla luovuta mitään käyttöoikeuksia sen patenti-, tavaramerkki-, tekijänoikeus- tai tapaoikeuksien nojalla eikä vastaavien kolmansien osapuolten oikeuksien nojalla.

Tässä kuvattuja tuotteita saa käyttää vain pätevä ja asianmukaisesti koulutettu henkilökunta noudattamalla täsmällisesti tässä asiakirjassa annettuja ohjeita, jotta tuotteiden asianmukainen ja turvallinen käyttö voidaan taata. Asiakirjan sisältö on luettava ja ymmärrettävä kokonaisuudessaan ennen näiden tuotteiden käyttöä.

MIKÄLI TÄSSÄ ANNETTUJA OHJEITA EI LUETA JA TÄSMÄLLISESTI NOUDATETA, SEURAUKSENA VOI OLLA TUOTTEIDEN VAURIOITUMINEN, HENKILÖVAHINKOJA JOKO KÄYTTÄJILLE TAI MUILLE JA MUITA OMAISUUSVAHINKOJA, MINKÄ LISÄKSI TUOTTEITA MAHDOLLISESTI KOSKEVAT TAKUUT MITÄTÖITYVÄT.

ILLUMINA EI OLE VASTUUSSA TÄSSÄ KUVATTUJEN TUOTTEIDEN VÄÄRINKÄYTÖSTÄ (MUKAAN LUKIEN TUOTTEEN OSAT JA OHJELMISTO).

© 2022 Illumina, Inc. Kaikki oikeudet pidätetään.

Kaikki tavaramerkit ovat Illumina, Inc:n tai niiden vastaavien omistajien omaisuutta. Tarkemmat tavaramerkkitiedot annetaan osoitteessa www.illumina.com/company/legal.html.

Sisällysluettelo

Versiohistoria	ii
Yleiskatsaus	1
Analyysimenetelmät	1
Ajon luominen	4
Asetukset	5
Analyysin tulokset	7
FASTQ-tiedostot	8
BAM-tiedostot	8
VCF-tiedostot	9
Analyysiraportin tarkastelu	14
Tekninen tuki	15

Yleiskatsaus

DRAGEN™ for Illumina® DNA Prep with Enrichment Dx -sovellus toteuttaa demultipleksoinnin, FASTQ-tiedostojen tuottamisen, lukukartoituksen ja kohdistuksen referenssigenomiin ja variantin tunnistuksen valitun analyysin työnkulun mukaisesti.

Analyysimenetelmät

DRAGEN for Illumina DNA Prep with Enrichment Dx toteuttaa demultipleksoinnin, FASTQ-tiedostojen tuottamisen, lukukartoituksen ja kohdistuksen viitegenomiin valitun työnkulun mukaisesti:

- FASTQ-tiedostojen tuottaminen
- Ituradan FASTQ- ja VCF-tiedostojen tuottaminen
- Somaattisen FASTQ:n ja VCF-tiedostojen tuottaminen

FASTQ-tiedoston luonti

Kootut sekvenssit kirjoitetaan FASTQ-tiedostoihin näytettä kohti. FASTQ-tiedostot ovat tekstitiedostoja, jotka sisältävät sekvensointitiedot ja laatuasteet vain yhdelle näytteelle. Erilliset FASTQ-tiedostot tuotetaan esimerkiksi virtauskyvetin kaistaa kohti, sekvensointilukua kohti. Näytteen nimi, joka annetaan ajon käyttöönoton aikana, sisältyy FASTQ-tiedoston nimeen. FASTQ-tiedostot ovat kohdistuksen pääasiallisia syötetietoja. FASTQ-tiedoston tuottamisen ensimmäinen vaihe on demultipleksointi. Demultipleksoinnissa nimetään kulterit, jotka läpäisevät suodattimen, näytteeseen vertailemalla kutakin indeksilukusekvenssiä ajoa varten nimettyihin indeksisekvensseihin. Tässä vaiheessa ei oteta huomioon laatuarvoja. Indeksireadit tunnistetaan suorittamalla seuraavat vaiheet:

- Näytteet numeroidaan alkaen 1:stä sen järjestyksen perusteella, jossa ne on lueteltu ajoa varten.
- Näyttenumero 0 on varattu klustereille, joita ei ole määrätty näytteeseen.
- Klusterit nimetään näytteeseen, kun indeksisekvenssi on tarkka osuma tai kun indeksilukua kohti on enintään yksi yhteensopimattomuus.

Ohjelmisto sisältää ORA-pakkauksen FASTQ-tiedostojen pakkaamiseksi. Kun käytetään ORA-muotoa (*.ora), md5-tarkistussumma säilytetään pakkaamisen ja pakkauksen purkamisen jakson jälkeen, jotta varmistetaan häviötön pakkaaminen.

DNA-kartoitus ja -kohdistus

Kartoituksen ensimmäinen vaihe on luoda alkuarvot lukua varten ja etsiä sitten tarkat osumat referenssigenomissa. Nämä tulokset jalostetaan sitten ajamalla täydet Smith-Waterman-kohdistukset sijaintipaikoissa, joissa on alkuarvo-osumien suurin tiheys. Tämä hyvin dokumentoitu algoritmi toimii vertailemalla luvun jokaista asemaa kaikkia referenssinehdokasemia vastaan. Nämä vertailut vastaavat mahdollisten kohdistusten matriisia luvun ja referenssin välillä. Smith-Waterman luo kullekin

näille ehdokaskohdistuksen asemille pisteytykset, joita käytetään sen arvioimiseen, saavuttaako paras kohdistus, joka kulkee kyseisen matriisisolun läpi, sen nukleotidiosuman tai -yhteensopimattomuuden (diagonaalinen liike), deleetoon (horisontaalinen liike) tai insertion (vertikaalinen liike) kautta. Osuma luvun ja referenssin välillä antaa bonuksen pisteytykseen ja yhteensopimattomuus tai indel-variantti tuo rangaistuksen. Korkeimman kokonaispisteytyksen polku matriisiin läpi on valittu kohdistus.

Pistemäärillä valitut tietyt arvot tässä algoritmossa osoittavat, miten tasapainotetaan kohdistukselle, jossa on useita mahdollisia tulkintoja, indel-variantin mahdollisuus yhtä tai useampaa SNP:tä vastaan tai etusija kohdistukselle ilman leikkautumista. DRAGEN-oletuspistearvot ovat käyttökelpoisia kohtalaisen pituisten lukujen kohdistamiseen koko referenssinä toimivaan ihmisgenomiin variantin tunnistussovelluksia varten. Mikä tahansa Smith-Waterman-pisteytyksen parametri edustaa genomisen mutaation ja sekvensointivirheiden epätarkkaa mallia. Eri tavalla säädetty kohdistuksen pistearvot voivat olla merkityksellisempiä joillekin sovelluksille.

DRAGEN, ituratavarianttien tunnistus

DRAGEN Germline Small Variant Caller -sovellus ottaa kartoitetut ja kohdistetut DNA-luvut syötteenä ja tunnistaa SNP:itä ja indel-variantteja sarakkeittaisen havaitsemisen ja haplotyyppien paikallisen *de novo* -kokoelman yhdistelmän kautta.

Tunnistettavissa olevat referenssialueet tunnistetaan ensin riittävällä kohdistuskattavuudella. Näiden referenssialueiden sisällä lajiteltujen lukujen nopea skannaus tunnistaa aktiiviset alueet, jotka ovat keskittyneet niiden kasaumasarakkeiden ympärille, joissa on näyttöä variantista. Aktiiviset alueet täytetään riittävällä kontekstilla, jotta katetaan viereinen merkittävä, ei-referenssin sisältämä sisältö. Jos esiintyy näyttöä indel-varianteista, aktiiviset alueet saavat lisätäyttöä.

Kohdistetut alueet katkaistaan kunkin aktiivisen alueen sisällä ja kootaan De Bruijn -kaavioksi. Katkaistujen lukujen reunat punnitaan havaintomäärittäin referenssisekvenssin muodostaessa rungon. Kun kaaviota on vähän puhdistettu ja yksinkertaistettu, eristetään kaikki lähteestä nieluun kulkevat polut ehdokashaplotyyppinä. Jokainen haplotyyppi on Smith-Waterman, joka on kohdistettu referenssigenomiin niiden varianttien tunnistamiseksi, joita se edustaa. Tätä tapahtumien sarjaa voidaan täydentää asemaan perustuvalla havaitsemisella. Kunkin luku-haplotyyppi-parin osalta todennäköisyys $P(r|H)$ kyseisen luvun havaitsemiseen, olettaen että haplotyyppi on todellinen aloittava näyte, arvioidaan käyttämällä paripiilotettua Markovin mallia (HMM).

Skannaamalla referenssiasemaa aktiivisen alueen päällä muodostetaan ehdokasgenotyypit varianttitapahtumien diploidisista yhdistelmistä (SNP:t tai indel-variantit). Kunkin tapahtuman (referenssi mukaan lukien) osalta ehdollinen todennäköisyys $P(r|e)$ kunkin limittäisen luvun havaitsemiseen arvioidaan $P(r|H)$ -maksimina niille haplotyypeille, jotka tukevat tapahtumaa. Nämä yhdistetään ehdolliseksi todennäköisyydeksi $P(r|e1e2)$ genotyypille (tapahtumapari) ja kerrotaan, jolloin saadaan ehdollinen todennäköisyys $P(R|e1e2)$ koko lukukasauman havaitsemiseen. Posteriorinen todennäköisyys $P(e1e2|R)$ kullekin diploidiselle genotyypille lasketaan Bayesin kaavaa käyttäen, ja voittaja tunnistetaan.

gVCF-mallissa, jota käytettiin skaalattavan moninäytevariantin tunnistukseen, DRAGEN Germline Small Variant Caller -sovellus voidaan ajaa näytettä kohti genomisen varianttitunnistuksen välitiedoston (gVCF) luomiseen. gVCF-tiedostoa voidaan sitten käyttää tehokkaaseen useiden näytteiden yhdistettyyn genotyyppiin määrittämiseen, mikä sallii näytteiden nopean kasvavan käsittelyn ja skaalauksen suuriin kohorttikokoihin.

Koska DRAGEN Germline Small Variant Caller -sovellus sisältää algoritmeja, jotka saavat sen kykenemään korreloitujen virheiden tehokkaaseen erottamiseen todellisista varianteista, suodatinsäännöt ovat erittäin yksinkertaisia.

DRAGEN, somaattisten varianttien tunnistus

DRAGEN Somatic Small Variant Caller -sovellus ottaa kartoitetut ja kohdistetut DNA-luvut syötteenä ja tunnistaa SNV:itä ja indel-variantteja haplotyyppien paikallisen *de novo* -kokoelman kautta aktiivisella alueella.

Tunnistettavissa olevat referenssialueet tunnistetaan ensin riittävällä kohdistuskattavuudella. Näiden referenssialueiden sisällä lajiteltujen lukujen nopea skannaus tunnistaa aktiiviset alueet, jotka ovat keskittyneet niiden kasaumasarakkeiden ympärille, joissa on näyttöä variantista tuumoriluvuissa. Aktiiviset alueet täytetään riittävällä kontekstilla, jotta katetaan viereinen merkittävä, ei-referenssin sisältämä sisältö. Jos esiintyy näyttöä indel-varianteista, aktiiviset alueet saavat lisätäyttöä.

Kohdistetut alueet katkaistaan kunkin aktiivisen alueen sisällä ja kootaan De Bruijn -kaavioksi. Katkaistujen lukujen reunat punnitaan havaintomäärittäin referenssisekvenssin muodostaessa rungon. Kun kaaviota on vähän puhdistettu ja yksinkertaistettu, eristetään kaikki lähteestä nieluun kulkevat polut ehdokashaplotyyppienä. Jokainen haplotyyppi on Smith-Waterman, joka on kohdistettu referenssigenomiin niiden varianttien tunnistamiseksi, joita se edustaa. Kunkin luku-haplotyyppi-parin osalta todennäköisyys $P(r|H)$ kyseisen luvun havaitsemiseen arvioidaan käyttämällä paripiilotettua Markovin mallia (HMM) olettaen, että haplotyyppi on todellinen aloittava näyte.

TLOD-pistemäärän määrittämiseksi DRAGEN Somatic Small Variant Caller -sovellus skannaa ensin referenssiasemittain kutakin ehdokastapahtumaa sekä referenssitapahtumaa aktiivisen alueen päällä. Ehdollinen todennäköisyys $P(r|e)$ kunkin limittäisen luvun havaitsemiseen arvioidaan $P(r|H)$ -maksimina niille haplotyypeille, jotka tukevat tapahtumaa. Nämä yhdistetään ehdolliseksi todennäköisyydeksi $P(r|E)$ tapahtumahypoteesille E, joka käsittää somaattisen referenssi- ja ehdokasalleelin seoksen mahdollisten alleelitiheyksien alueesta, ja kerrotaan, jolloin saadaan ehdollinen todennäköisyys $P(R|E)$ koko lukukasauman havaitsemiseen. Tästä lasketaan TLOD-pistemäärä näyttönä, että jokin ALT-alleeli esiintyy kasvainnäytteessä tietyssä sijainnissa.

Ajon luominen

Käytä seuraavia vaiheita ajon ottamiseksi käyttöön Illumina Run Manager -laitteessa joko NovaSeq 6000Dx -laitteella tai käyttäen verkkotietokoneen selainta. Näytetiedot voidaan syöttää manuaalisesti tai tuomalla näytetiedoston.

Sovelluksen ja ajon asetukset

1. Valitse Runs (Ajot) -näytöstä **Create Run** (Luo ajo).
2. Valitse DRAGEN for Illumina DNA Prep with Enrichment Dx -sovellus ja sitten **Next** (Seuraava).
3. Anna Run Settings (Ajon asetukset) -näytössä ajon nimi. Analyysin aikana ajo tunnustetaan sekvensoinnista ajon nimen perusteella.
4. **[Valinnainen]** Syötä ajon kuvaus ajon tunnistamisen helpottamiseksi.
5. Varmista, että valittu Library Prep Kit (Kirjaston valmistepakkaus) on Illumina DNA Prep with Enrichment Dx -kirjaston valmistepakkaus.
6. Valitse haluttu indeksiadapteripakkaus.
7. Anna Read Length (lukupituus).
Read 1 (Luku 1):n ja Read 2 (Luku 2):n oletusarvo on 151 jaksoa.
Index 1 (Indeksi 1):n ja Index 2 (Indeksi 2):n oletusarvo on 10 jaksoa.
8. **[Valinnainen]** Anna kirjastoputken tunnus.
9. Valitse **Next** (Seuraava).

Näytetiedot

Käytä Sample Data (Näytetiedot) -näytön taulukkoa näytteen tietojen syöttämiseen manuaalisesti. Valitse vaihtoehtoisesti **Import Samples** (Tuo näytteet) näytetietojen lataamiseksi. Ks. tietoja näytetietojen tuomisesta: [Import Samples \(Tuo näytteet\) sivulla 5](#).

Näytteiden syöttäminen manuaalisesti

1. Anna ainutkertainen näytetunnus Sample ID (Näytetunnus) -kenttään.
2. Käytä kohtaa **Plate - Well Position** (Levy - Kuopan asema) kuopan aseman valitsemiseen. i7 Index, Index 1, i5 Index ja Index 2 -indeksikentät täyttyvät automaattisesti.
3. **[Valinnainen]** Anna kirjaston nimi.
4. Lisää rivejä ja toista vaiheita 1–3 tarpeen mukaan, kunnes kaikki näytteet on lisätty taulukkoon.
5. Valitse **Next** (Seuraava).

Import Samples (Tuo näytteet)

Sample Data (Näytetiedot) -näytössä käytettävissä on malli (*.csv) lataamista varten, kun suunnitellaan ajoja Illumina Run Manager -laitteessa käyttäen verkkotietokoneen selainta.

1. Valitse **Download Template** (Lataa malli) ladataksesi tyhjän CSV-tiedoston.
2. Anna näytetiedot kyseisestä CSV-tiedostosta ja tallenna tiedosto.
Näytetiedoston CVS-tiedosto sisältää seuraavat tietosarakkeet: Sample ID (Näytetunnus), Plate - Well Position (Levy - Kuopan asema), **Optional** Library Name (Valinnainen kirjaston nimi).
3. Valitse **Import Samples** (Tuo näytteet) ja selaa CSV-tiedoston sijaintiin.
4. Valitse **Next** (Seuraava).

Analyysiasetukset

1. Valitse haluttu analyysityönkulku:
 - FASTQ-tiedostojen tuottaminen
 - Ituradan FASTQ- ja VCF-tiedostojen tuottaminen ituratan työnkululle
 - Somaattisten FASTQ:n ja VCF-tiedostojen tuottaminen somaattiselle työnkululle
2. **[Valinnainen]** Valitse tarvittaessa **Generate ORA compressed FASTQs** (Luo ORA-pakatut FASTQ:t) -valintaruutu, niin FASTQ ORA -pakkaaminen otetaan käyttöön.
3. **[VCF-tiedostojen tuottamisen työnkulut]** Käytä avattavaa **Manifest File Selection** (Manifestitiedoston valinta) -valikkoa jonkin manifestitiedoston valitsemiseen.
Manifestitiedosto on tarvittava syöte DRAGEN for Illumina DNA Prep with Enrichment Dx -laitteelle. Manifesti on sarkaineroteltu BED-tiedosto (*.bed), joka määrittää kohdistettujen referenssialueiden nimet ja sijainnit.
4. **[Somatic FASTQ- VCF-tiedostojen tuottamisen työnkulku]** Valitse kohinatiedosto **Noise File Selection** (Kohinatiedoston valinta) -pudotusvalikosta.
BED-tiedosto, jossa on käyttöpaikkakohtainen kohinataso, voidaan määrittää systemaattisen kohinan poissuodattamista varten. Katso lisätietoja: [Kohinan suodattaminen sivulla 6](#).
5. Valitse **Next** (Seuraava).

Ajo Tarkastelu

1. Tarkastele Run Settings (Ajoasetukset)-, Sample Data (Näytetiedot)- ja Analysis Settings (Analyysiasetukset) -näytöissä annettuja tietoja Review (Tarkastelu) -näytöllä.
2. Valitse **Save** (Tallenna).
Ajo tallennetaan Runs (Ajot) -näytön Planned (Suunnitellut) -välilehdelle.

Asetukset

Valitse sovellus Applications (Sovellukset) näytöstä, jotta voit katsella tämänhetkisiä asetuksia ja muuttaa asetuksia.

Määrittäminen

Määrittämisnäyttö näyttää seuraavat sovellusasetukset:

- **Library Prep Kits** (Kirjaston valmistelupakkaukset) – Näyttää kirjaston oletusarvoisen valmistelupakkauksen kyseistä sovellusta varten. Tätä asetusta ei voi muuttaa.
- **Index Adapter Kits** (Indeksiadapteripakkaukset) – Näyttää oletusarvoisen indeksiadapteripakkauksen kyseistä sovellusta varten. Tätä asetusta ei voi muuttaa.
- **Lukupituudet** – Lukupituudet on asetettu sovellusta varten oletuksena arvoon 151, mutta sitä voidaan muuttaa ajon luomisen aikana.
- **Manifesti- ja kohinatiedostot** – Lataa ja vaihda asetukset manifesti- ja kohinatiedostoja varten.
 - Valitse **Upload File** (Lataa tiedosto) ladataksesi tiedostoja analyysissä käytettäviksi.
 - Valitse **Default** (Oletus) -valintapainike asettaaksesi manifesti- tai kohinatiedoston oletukseksi, joka valitaan ajon luomisen aikana, kun sovellus valitaan.
 - Valitsemalla **Enabled** (Käytössä) -valintaruudun voit asettaa tiedoston näkymään avattavassa valikossa ajon luomisen aikana.

Luvat

Käytä Permissions (Luvat) -näytön valintaruutuja sovelluksen käyttöoikeuksien hallintaan.

Kohinan suodattaminen

Kun käytetään somaattista työkulkua, käytettävissä on systemaattinen kohinan suodatus. Kyseistä suodatinta voi käyttää Tumor-Normal (Kasvain-Normaali) -tilassa, mutta se on erityisen hyödyllinen Tumor-Only (Vain kasvain) -ajoissa, kun kaltaistettua normaalia ei ole käytettävissä.

Normaalinäytteistä on luotava systemaattisen kohinan BED-tiedosto. On suositeltavaa rakentaa systemaattiset kohinatiedostot, jotka ovat kirjastovalmistelle, sekvensointijärjestelmälle ja paneelille spesifisiä. On suositeltavaa käyttää noin 50:tä normaalinäytettä kohinatiedoston luomiseen.

Analyysin tulokset

DRAGEN for Illumina DNA Prep with Enrichment Dx tallentaa seuraavat tiedot analyysikansioon. Vain iturata- ja somaattiset työkulut aikaansaavat PDF-tiedoston.

- Käytetty manifestitiedosto
- Ohjelmistoversio
- Näytetunnukset
- Kohdistettuja lukuja yhteensä
- Kohdistettujen lukujen prosenttiosuus näytettä kohti
- Tunnistettujen SNV:iden määrä näytettä kohti
- Tunnistettujen indel-varianttien määrä näytettä kohti
- Kattavuuden tilastotiedot

Analyysin tuotostiedostot

Sovellus tuottaa seuraavat tulostetiedostot. Tarkat tuotetut tiedostot ovat käytettävän analyysityönkulun mukaisia. Tulostetiedostot sijaitsevat analyysikansiossa.

Tulostetiedosto	Kuvaus
FASTQ (*.fastq.gz tai *.fastq.ora)	Välitiedostot, jotka sisältävät laatupisteytykseen perustuvia emästen tunnistuksia. FASTQ-tiedostot ovat kohdistusvaiheen pääasiallisia syötetietoja. Jos valitaan ORA-pakkaus, tiedoston nimi heijastaa tätä.
Kohdistuksen BAM-tiedostot (*.bam)	Sisältää tietyn näytteen kohdistettuja readeja.
Genoomiset VCF-tiedostot (*.gvct.gz)	Sisältää kunkin aseman genotyypin riippumatta siitä, tunnistetaanko ne varianteiksi vai viitteiksi.
VCF-tiedostot (*.vcf.gz)	Sisältää variantteja, jotka on tunnistettu kustakin asemasta.
Ajon metriikkaraportti (*.csv)	Sisältää laatumetriikan ajosta, myös kokonaissaannon ja Q30-pistemäärät.

FASTQ-tiedostot

FASTQ (*.fastq.gz, *.fastq.ora) on tekstipohjainen tiedostomuoto, joka sisältää emästunnistukset ja laatuarvot lukua kohti. Kukin tiedosto sisältää seuraavat tiedot:

- Näytteen tunniste
- Sekvenssi
- Plus-merkki (+)
- Phred-laatuasteikot ASCII + 33 -koodatussa muodossa

Näytteen tunnisteiden muoto on seuraava.

```
@Instrumentti:Ajotunniste:Virtauskyvettitunniste:Kaista:Ruutu:X:Y Read-
numero:Suodatinlippu:0:Näyttenumero
Esimerkki:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAA9#:<#<;<<<????#=#
```

BAM-tiedostot

BAM-tiedosto (*.bam) on pakattu binääri-versio SAM-tiedostosta (sekvenssien kohdistuskartta), jota käytetään esittämään kohdistettuja sekvenssejä enintään 128 Mb:een saakka. BAM-tiedostoissa käytetään tiedostojen nimeämismuotoa `sampleName_S#.bam`. # on näytteen numero määritettynä järjestyksessä, jossa näytteet luetellaan ajoa varten. Monisolmuksessa tilassa S# on asetettu arvoon S1 huolimatta näytejärjestyksestä.

BAM-tiedostot sisältävät otsikko-osion ja kohdistusosion:

- Otsikko – Sisältää tietoa koko tiedostosta, kuten nimen, näytteen pituuden ja kohdistusmenetelmän. Kohdistusosion kohdistukset liittyvät otsikko-osion tiettyihin tietoihin.
- Kohdistukset – Sisältää readin nimen, sekvenssin, laadun, kohdistustiedot ja mukautetut tunnisteet. Readin nimeen sisältyy kromosomi, aloituskoordinaatti, kohdistuslaatu ja osuman kuvausmerkkijono.

Kohdistusosioon sisältyvät seuraavat tiedot kutakin readia tai read-paria:

- AS: Paired-end-kohdistuksen laatu.
- RG: Luvun ryhmä, joka osoittaa tietyn näytteen lukujen määrän.
- BC: Viivakooditunniste, joka osoittaa lukuun liittyvän demultipleksoidun näytteen tunnuksen.
- SM: Yhden pään kohdistuksen laatu.
- XC: Osuman kuvaajan merkkijono.

- XN: Amplikoninimitunniste, joka tallentaa luvun BAM-indeksitiedostoihin (*.bam.bai) liittyvän amplikonin tunnuksen anna vastaavan BAM-tiedoston indeksi.

VCF-tiedostot

Variantin tunnistusmuodon (*.vcf) tiedostot sisältävät tietoa referenssigenomin tietyistä asemista havaituista varianteista.

VCF-tiedoston otsikkoon sisältyy VCF-tiedostomuotoversio ja varianttitunnistinversio, ja siinä luetellaan tiedoston muissa osissa käytetyt merkinnät. VCF-otsikkoon sisältyy myös viitegenomitiedosto ja BAM-tiedosto. Otsikon viimeinen rivi sisältää tietorivien sarakkeiden otsakkeet. Jokainen VCF-tiedoston tietoriveistä sisältää tietoa yhdestä variantista.

Taulukko 1 VCF-tiedostojen otsakkeet

Otsake	Kuvaus
CHROM	Viitegenomin kromosomi. Kromosomit tulevat näkyviin samassa järjestyksessä kuin referenssi-FASTQ-tiedosto.
POS	Variantin yksittäisen emäksen asema viitekromosomissa. Yhden nukleotidin varianteille (SNV:t) tämä asema on referenssiemäs variantille. Indel-varianteille tämä asema on referenssiemäs, joka välittömästi edeltää varianttia.
ID (Tunnus)	rs (referenssi-SNP) -numero dbSNP.txt-tiedostosta saadulle SNP:lle. Jos tässä sijainnissa on useita rs-numeroita, luettelo erotellaan puolipisteillä. Jos tässä asemassa ei ole dbSNP-tietuetta, käytetään puuttuvan arvon merkkiä ('.').
REF (VIITE)	Viitegenotyyppi. Esimerkiksi yksittäisen T:n poisto esitetään viitteenä TT ja vuorotellen T:nä. Yksittäinen A - T-nukleotidivariantti esitetään viitteenä A ja vuorotellen T:nä.
ALT	Viitereadista poikkeavat alleelit. Esimerkiksi yksittäisen T:n lisäys esitetään viitteenä A ja vuorotellen AT:nä. Yksittäinen A - T-nukleotidivariantti esitetään viitteenä A ja vuorotellen T:nä.
QUAL	Varianttitunnistimen määrittämä Phred-skaalattu laatupisteytys. Korkeammat pistemäärät viittaavat variantin suurempaan luottamukseen ja alhaisempaan virheiden todennäköisyyteen. Laatupistemäärän Q kohdalla arvioitu virheen todennäköisyys on 10 ^{-Q/10} . Esimerkiksi Q30-tunnistusten joukon virheaste on 0,1 % . Monet varianttitunnistimet määrittävät laatupistemääriä tilastomalliensa perusteella. Ne ovat korkealla tasolla havaitun virheasteen suhteen.

Taulukko 2 VCF-tiedostomerkinnot

Otsake	Kuvaus
FILTER (SUODATIN)	<p>Mikäli kaikki suodattimet hyväksytään, suodatinsarakkeeseen kirjoitetaan PASS (HYVÄKSYTTY).</p> <p>Ituradan työnkulun mahdollisia FILTER (SUODATIN) -tietueita ovat mm. seuraavat:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL – Käytetään, jos SNP-variantin QUAL-pistemäärä ei saavuta kynnysarvoa • DRAGENIndelHardQUAL – Käytetään, jos indel-variantin QUAL-pistemäärä ei saavuta kynnysarvoa • LowDepth – Sijainti suodatetaan, koska kattavuuden syvyys ei saavuta kynnysarvoa • LowGQ – Sijainti suodatetaan, koska genotyypin laatu ei saavuta kynnysarvoa • PloidyConflict – Genotyypitunnistus varianttitunnistuksesta ei ole yhteneväinen kromosomin ploidian kanssa • base_quality – Sijainti suodatetaan, koska alt-lukujen emäksen mediaanilaatu tässä sijainnissa ei saavuta kynnysarvoa • filtered_reads – Sijainti suodatetaan, koska liian suuri lukujen osuus on poissuodatettu • fragment_length – Sijainti suodatetaan, koska alt-lukujen fragmentin mediaanipituuden ja ref-lukujen fragmenttipituuden välinen absoluuttinen erotus tässä sijainnissa ylittää kynnysarvon • low_depth – Sijainti suodatetaan, koska lukusyvyys on liian pieni • low_frac_info_reads – Sijainti suodatetaan, koska informatiivisten lukujen osuus alittaa kynnysarvon • low_normal_depth – Sijainti suodatetaan, koska normaalin näytteen lukusyvyys on liian pieni • long_indel – Sijainti suodatetaan, koska indel-pituus on liian pitkä • mapping_quality – Sijainti suodatetaan, koska alt-lukujen kartoituksen mediaanilaatu tässä sijainnissa ei saavuta kynnysarvoa • multiallelic – Sijainti suodatetaan, koska enemmän kuin kaksi alt-alleelia läpäisee kasvain-LOD:n • non_homref_normal – Sijainti suodatetaan, koska normaalin näytteen genotyyppi ei ole homotsygoottinen referenssi • no_reliable_supporting_read – Sijainti suodatetaan, koska ei ole luotettavaa tukevaa somaattista lukua • panel_of_normals – Nähdään vähintään yhdessä näytteessä normaalien vcf-paneelissa • read_position – Sijainti suodatetaan, koska luvun ja tämän sijainnin aloituksen/päätymisen välisten etäisyyksien mediaani alittaa kynnysarvon • RMxNRepeatRegion – Sijainti suodatetaan, koska kaikki tai osa varianttialleelista on referenssin toistoa • strand_artifact – Sijainti suodatetaan, koska juosteessa on vakava systemaattinen virhe

Otsake	Kuvaus
FILTER (SUODATIN) (jatkuu)	<ul style="list-style-type: none"> • str_contraction – Sijainti suodatetaan, koska epäillään PCR-virhettä, jossa alt-alleeli on yhden toistoyksikön verran pienempi kuin referenssi • too_few_supporting_reads – Sijainti suodatetaan, koska on liian vähän tukevia lukuja kasvainnäytteessä • weak_evidence – Somaattisen variantin pistemäärä ei saavuta kynnsarvoa Somaattisen työnkulun mahdollisia FILTER (SUODATIN) -tietueita ovat mm. seuraavat: <ul style="list-style-type: none"> • base_quality – Sijainti suodatetaan, koska alt-lukujen emäksen mediaanilaatu tässä sijainnissa ei saavuta kynnsarvoa • filtered_reads – Sijainti suodatetaan, koska liian suuri lukujen osuus on poissuodatettu • fragment_length – Sijainti suodatetaan, koska alt-lukujen fragmentin mediaanipituuden ja ref-lukujen fragmenttipituuden välinen absoluuttinen erotus tässä sijainnissa ylittää kynnsarvon • low_depth – Sijainti suodatetaan, koska lukusyvyys on liian pieni • low_frac_info_reads – Sijainti suodatetaan, koska informatiivisten lukujen osuus alittaa kynnsarvon • low_normal_depth – Sijainti suodatetaan, koska normaalin näytteen lukusyvyys on liian pieni • long_indel – Sijainti suodatetaan, koska indel-pituus on liian pitkä • mapping_quality – Sijainti suodatetaan, koska alt-lukujen kartoituksen mediaanilaatu tässä sijainnissa ei saavuta kynnsarvoa • multiallelic – Sijainti suodatetaan, koska enemmän kuin kaksi alt-alleelia läpäisee kasvain-LOD:n • non_homref_normal – Sijainti suodatetaan, koska normaalin näytteen genotyyppi ei ole homotsygoottinen referenssi • no_reliable_supporting_read – Sijainti suodatetaan, koska ei ole luotettavaa tukevaa somaattista lukua • panel_of_normals – Nähdään vähintään yhdessä näytteessä normaalien vcf-paneelissa • read_position – Sijainti suodatetaan, koska luvun ja tämän sijainnin aloituksen/päättymisen välisten etäisyyksien mediaani alittaa kynnsarvon • RMxNRepeatRegion – Sijainti suodatetaan, koska kaikki tai osa varianttialleelista on referenssin toistoa • strand_artifact – Sijainti suodatetaan, koska juosteessa on vakava systemaattinen virhe • str_contraction – Sijainti suodatetaan, koska epäillään PCR-virhettä, jossa alt-alleeli on yhden toistoyksikön verran pienempi kuin referenssi • too_few_supporting_reads – Sijainti suodatetaan, koska on liian vähän tukevia lukuja kasvainnäytteessä • weak_evidence – Somaattisen variantin pistemäärä ei saavuta kynnsarvoa • systematic_noise – Sijainti suodatetaan, koska on näyttöä systemaattisesta kohinasta normaaleissa

Otsake	Kuvaus
INFO (TIEDOT)	<p>Ituradan työnkulun mahdollisia INFO (TIEDOT) -tietueita ovat mm. seuraavat:</p> <ul style="list-style-type: none"> • AC – Genotyypeissä oleva alleelien määrä kunkin ALT-alleelin osalta luetellussa järjestyksessä. • AF – Alleelin taajuus kunkin ALT-alleelin osalta luetellussa järjestyksessä. • AN – Alleelien kokonaismäärä tunnistetuissa genotyypeissä. • DB – dbSNP-jäsenyys. • FS – Phred-skaalattu p-arvo Fisherin tarkalla testillä juosteen systemaattisen virheen havaitsemiseksi. • QD – Variantin luottamus/laatu syvyyttä kohti. • R2_5P_bias – Pistemäärä, joka perustuu parin systemaattiseen virheeseen ja etäisyyteen 5'-päästä. • SOR – 2x2-kontingenssitaulukon symmetrinen kerroinsuhde juosteen vinouman havaitsemiseksi. • DP – Likimääräinen lukusyvyys (informatiivinen ja ei-informatiivinen); jotkin luvut on ehkä suodatettu mapq:n perusteella jne. • END – Välin päättymisasema. • FractionInformativeReads – Informatiivisten lukujen osuus kokonaisluvuista. • MQ – RMS-kartoituksen laatu. • MQRankSum – Z-pistemäärä Alt vs. Ref -lukukartoituksen laatujen Wilcoxonin U-testistä. • ReadPosRankSum – Z-pistemäärä Alt vs. Ref -lukukartoituksen aseman systemaattisen virheen Wilcoxonin U-testistä. • SOMATIC – Vähintään yksi variantti tässä asemassa on somaattinen. <p>Somaattisen työnkulun mahdollisia INFO (TIEDOT) -tietueita ovat mm. seuraavat:</p> <ul style="list-style-type: none"> • DP – Likimääräinen lukusyvyys (informatiivinen ja ei-informatiivinen); jotkin luvut on ehkä suodatettu mapq:n perusteella jne. • END – Välin päättymisasema. • FractionInformativeReads – Informatiivisten lukujen osuus kokonaisluvuista. • MQ – RMS-kartoituksen laatu. • MQRankSum – Z-pistemäärä Alt vs. Ref -lukukartoituksen laatujen Wilcoxonin U-testistä. • ReadPosRankSum – Z-pistemäärä Alt vs. Ref -lukukartoituksen aseman systemaattisen virheen Wilcoxonin U-testistä. • AQ – Systemaattisen kohinan pistemäärä. • hotspot – Tunnettu somaattinen sijainti, käytetään luottamuksen kasvattamiseksi tunnistuksessa. • SOMATIC – Vähintään yksi variantti tässä asemassa on somaattinen.

Otsake	Kuvaus
FORMAT (MUOTO)	<p>Muotosarakkeessa luetellaan kaksoispisteillä erotettuja kenttiä. Esimerkiksi GT:GQ.</p> <p>Ituradan työnkulun käytettävissä olevia kenttiä ovat mm. seuraavat:</p> <ul style="list-style-type: none"> • AD – Alleelisyvyydet (laskien vain informatiiviset luvut kokonaisluvuista) ref- ja alt-alleeleille luetellussa järjestyksessä. • AF – Alleeliosuudet alt-alleeleille luetellussa järjestyksessä. • DP – Likimääräinen lukusyvyys (luvut, joissa MQ=255 tai huonoja pareja, suodatetaan). • F1R2 – Lukujen määrä F1R2-parin suunnassa, jotka tukevat jokaista alleelia. • F2R1 – Lukujen määrä F2R1-parin suunnassa, jotka tukevat jokaista alleelia. • GP – Phred-skaalatut jälkitodennäköisyydet genotyypeille, jotka määrittellään VCF-tiedoston tiedoissa. • GQ – Genotyypin laatu. • GT – Genotyyppi. 0 vastaa viite-emästä, 1 vastaa ALT-sarakkeen ensimmäistä merkintää jne. Eteenpäin suuntautuva vinoviiva (/) osoittaa, että vaiheistustietoja ei ole saatavilla. • MB – Näytekohtaisen komponentin tilastotiedot parin systemaattisen virheen havaitsemiseksi. • PL – Normalisoidut, Phred-skaalatut todennäköisyydet genotyypeille, jotka määrittellään VCF-tiedoston tiedoissa. • PRI – Phred-skaalatut aiemmat todennäköisyydet genotyypeille. • PS – Fysikaalisen tahdistuksen tunnustiedot, jossa jokainen ainutkertainen tunnus tietyn näytteen sisällä (mutta ei näytteiden poikki) liittyy tietueet tiettyyn tahdistusryhmään. • SB – Näytekohtaisen komponentin tilastotiedot, jotka käsittävät Fisherin tarkan testin juosteen systemaattisen virheen havaitsemiseksi. • SQ – Somaattinen laatu. <p>Somaattisen työnkulun käytettävissä olevia kenttiä ovat mm. seuraavat:</p> <ul style="list-style-type: none"> • AD – Alleelisyvyydet (laskien vain informatiiviset luvut kokonaisluvuista) ref- ja alt-alleeleille luetellussa järjestyksessä. • AF – Alleeliosuudet alt-alleeleille luetellussa järjestyksessä. • DP – Likimääräinen lukusyvyys (luvut, joissa MQ=255 tai huonoja pareja, suodatetaan). • F1R2 – Lukujen määrä F1R2-parin suunnassa, jotka tukevat jokaista alleelia. • F2R1 – Lukujen määrä F2R1-parin suunnassa, jotka tukevat jokaista alleelia. • GT – Genotyyppi. 0 vastaa viite-emästä, 1 vastaa ALT-sarakkeen ensimmäistä merkintää jne. Eteenpäin suuntautuva vinoviiva (/) osoittaa, että vaiheistustietoja ei ole saatavilla. • MB – Näytekohtaisen komponentin tilastotiedot parin systemaattisen virheen havaitsemiseksi. • PS – Fysikaalisen tahdistuksen tunnustiedot, jossa jokainen ainutkertainen tunnus tietyn näytteen sisällä (mutta ei näytteiden poikki) liittyy tietueet tiettyyn tahdistusryhmään.

Otsake	Kuvaus
FORMAT (MUOTO) (jatkuu)	<ul style="list-style-type: none"> • SB – Näytekohtaisen komponentin tilastotiedot, jotka käsittävät Fisherin tarkan testin juosteen systemaattisen virheen havaitsemiseksi. • SQ – Somaattinen laatu.
SAMPLE (NÄYTE)	Näytesarakkeessa ilmoitetaan FORMAT (MUOTO) -sarakkeessa määritetyt arvot.

Genomin VCF-tiedostot

Genomin VCF (*.gvcf.gz) -tiedostot noudattavat joukkoa käytänteitä genomin sisällä olevien kaikkien sijaintien edustamiseksi kohtuullisen kompaktissa muodossa. gVCF-tiedostoihin sisältyvät kohdealueen kaikki sijainnit yhdessä tiedostossa kullekin näytteelle. gVCF-tiedostossa esitetään ”ei tunnistuksia” asemissa, jotka eivät läpäise kaikkia suodattimia. Genotyypin (GT) tunnistella ./.. viitataan Ei tunnistuksia -tilaan.

Analyysiraportin tarkastelu

Tällä hetkellä käynnissä olevat ajot näytetään Active (Aktiivinen) -välilehdellä. Päätyneet ajot näytetään Completed (Päätyneet) -välilehdellä. Ks. lisätietoja tulosten tarkastelusta: [NovaSeq 6000Dx - tuotedokumentaatio \(asiakirja nro 200010105\)](#).

Tekninen tuki

Teknisissä kysymyksissä voit ottaa yhteyttä Illuminan tekniseen tukeen.

Verkkosivusto: www.illumina.com

Sähköposti: techsupport@illumina.com

Illuminan teknisen tuen puhelinnumerot

Alue	Ilmainen	Kansainvälinen
Australia	+61 1800 775 688	
Itävalta	+43 800 006249	+43 1 9286540
Belgia	+32 800 77 160	+32 3 400 29 73
Kanada	+1 800 809 4566	
Kiina		+86 400 066 5835
Tanska	+45 80 82 01 83	+45 89 87 11 56
Suomi	+358 800 918 363	+358 9 7479 0110
Ranska	+33 8 05 10 21 93	+33 1 70 77 04 46
Saksa	+49 800 101 4940	+49 89 3803 5677
Hongkong, Kiina	+852 800 960 230	
Intia	+91 8006500375	
Indonesia		0078036510048
Irlanti	+353 1800 936608	+353 1 695 0506
Italia	+39 800 985513	+39 236003759
Japani	+81 0800 111 5011	
Malesia	+60 1800 80 6789	
Alankomaat	+31 800 022 2493	+31 20 713 2960
Uusi-Seelanti	+64 800 451 650	
Norja	+47 800 16 836	+47 21 93 96 93
Filippiinit	+63 180016510798	
Singapore	1 800 5792 745	
Etelä-Korea	+82 80 234 5300	
Espanja	+34 800 300 143	+34 911 899 417

Alue	Ilmainen	Kansainvälinen
Ruotsi	+46 2 00883979	+46 8 50619671
Sveitsi	+41 800 200 442	+41 56 580 00 00
Taiwan, Kiina	+886 8 06651752	
Thaimaa	+66 1800 011 304	
Yhdistynyt kuningaskunta	+44 800 012 6019	+44 20 7305 7197
Yhdysvallat	+1 800 809 4566	+1 858 202 4566
Vietnam	+84 1206 5263	

Käyttöturvallisuustiedotteet (KTT) – saatavilla Illuminan verkkosivustolta osoitteesta support.illumina.com/sds.html.

Tuotedokumentaatio – ladattavissa osoitteesta support.illumina.com.



Illumina
5200 Illumina Way
San Diego, California 92122 U.S.A.
+1 800 809.ILMN (4566)
+1 858 202 4566 (Pohjois-Amerikan
ulkopuolella)
techsupport@illumina.com
www.illumina.com

CE



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

Rahoittaja Australiassa

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

IN VITRO -DIAGNOSTISEEN KÄYTTÖÖN

© 2022 Illumina, Inc. Kaikki oikeudet pidätetään.

illumina[®]