

illumina®

TruSight Whole Genome Analysis Application

Tuotedokumentaatio

ILLUMINAN OMISTAMAA SISÄLTÖÄ
Asiakirjanro 200049931 v00
Huhtikuu 2024
DIAGNOSTISEEN IN VITRO -KÄYTTÖÖN.

Versiohistoria

Asiakirja	Päivämäärä	Muutoksen kuvaus
Asiakirja nro 200049931 v00	Huhtikuu 2024	Ensimmäinen versio.

Tämä asiakirja ja sen sisältö ovat Illumina, Inc:n ja sen tytäryhtiöiden ("Illumina") omaisuutta, ja ne on tarkoitettu ainoastaan Illumina-yhtiön 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 Illumina-yhtiöltä ennakkoon saatua kirjallista lupaa. Illumina ei tällä asiakirjalla luovuta mitään käyttöoikeuksia sen patentti-, 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 KAIKKIA 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).

© 2024 Illumina, Inc. Kaikki oikeudet pidätetään.

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

Sisällysluettelo

Versiohistoria	ii
Yleiskatsaus	1
Aloittaminen	1
Asetukset	2
Ajon luominen	4
Suorita revisio	5
Analyysin asettaminen uudelleen jonoon	6
Ajon ja tulosten tarkasteleminen	7
Tuotostiedoston yhteenveto	8
Näyteraportin tiedot	8
Variantin kutsutiedoston tiedot	12
FASTQ-tiedostot	20
CRAM-tiedostot	21
Tekninen tuki	22

Yleiskatsaus

TruSight Whole Genome Analysis Application -ohjelmistoa käytetään sekvensointiajajien suunnitteluun -laitteelle TruSight Whole Genome ja analyysin automaattiseen aloittamiseen ajon päätyttyä. Analyysiin kuuluu aineenpurkaminen, FASTQ-luonti, lukukartoitus, linjaus kuvaajaa käyttävän GrCh38/hg38-ihmisen viitegenomin kanssa ja varianttien tunnistusta käyttäen -Illumina DRAGEN Server koodia NovaSeq 6000Dx.

Analyysin työnkulun eri vaiheissa sovellus suorittaa laadunvalvonnan (QC) määritettyjen sekvensointi-, FASTQ- ja näytekirjastometriikoiden mukaisesti ja luo tulokset sisältäviä raportteja. Jos näyte läpäisee kaikki laadunvalvontavaiheet, sovellus luo tukitiedostoja käytettäväksi alavirran ituradan sovelluksissa.

TruSight Whole Genome Analysis Application suorittaa DRAGEN varianttitunnistajat, mukaan lukien Pienten varianttien tunnistin, Kopiolukuvaihteluiden (CNV) tunnistin ja Toistolaaajennuksen havaitseminen ExpansionHunter-ohjelmistolla.

Sovellus suorittaa myös pienten varianttien matalan, keskitason tai korkean luotettavuustason annotaation ja sisällyttää tämän annotaation tulostiedostoon.

Aloittaminen

Varmista, että TruSight Whole Genome Analysis Application on asennettu NovaSeq 6000Dx laitteeseen, jota käytetään sekvensointiin osana TruSight Whole Genome. Asennetut sovellukset löytyvät NovaSeq 6000Dx Instrument:n Sovellukset-näytöltä tai Illumina Run Manager stä verkkotietokoneen selaimella. Jos tarvitset apua asennuksen aikatauluttamisessa, ota yhteys paikalliseen Illumina kenttäedustajaan.

Varastointivaatimus

Ks. NovaSeq 6000Dx -tuotedokumentaatio (asiakirja nro 200010105)- ja DRAGEN Server for NovaSeq 6000Dx, tuotedokumentaatio (asiakirjanro 200014171)-kohdista yleisiä tietoja tietojen antamisesta ja tallentamisesta.

TruSight Whole Genome Analysis Application tuottaa tiedot suorituskansioon ja analyysikansioon ulkoisessa tallennustilassa. Minimitallennusvaatimukset voidaan määrittää likimääräisesti kuhunkin kansioon lähetetyn datan koosta alla esitettyä yksittäistä sekvensointiajajoa varten.

Määrittäminen	Suorita kansio (Gt)	Analyysikansio (Gt)
S2-virtaussolu (6 näytettä)	~430	~350
S4-virtaussolu (16 näytettä)	~1110	~890

Arvioitu analyysiaika

Analyysi käynnistyy automaattisesti, kun sekvensointiajo on valmis, ja se tapahtuu peräkkäin ajon sisältämissä näytteissä. Tietojen tuotostiedostot ovat käytettävissä ulkoisessa tallennustilassa, kun analyysi on valmis kaikille ajon näytteille ja kopiosiiro ulkoiseen tallennustilaan on valmis. Kun sekvensointiajo aloitetaan sekä A- että B-puolella samanaikaisesti, sekvensointi suoritetaan samanaikaisesti. TruSight Whole Genome Analysis Application analysoi nämä sekvensointijaksot peräkkäin sekvensoinnin päätyttyä. Sekvensoinnin ja siirron ensin suorittanut ajo analysoidaan ensin. Toinen sekvensointiajo siirretään ja asetetaan jonoon analysointia varten, kun ensimmäinen analyysi on valmis. Katso kohdasta [Ajon ja tulosten tarkasteleminen sivulla 7](#), miten voit määrittää aktiivisten tai epäonnistuneiden ajojen tilan.

Arvioitu aika, kunnes analyysitulokset ovat saatavilla sekvensoinnin jälkeen, näytetään alla tilanteessa, jossa A- ja B-puoli ladataan samanaikaisesti samalla määrityksellä.

Määrittys	Analyysiajo 1 (tuntia)	Analyysiajo 2 (tuntia)
S2-virtaussolu (6 näytettä)	~12	~24
S4-virtaussolu (16 näytettä)	~24	~48

Asetukset

Valitse TruSight Whole Genome Analysis Application Sovellukset-näytössä nähdäksesi nykyiset määrittymiset ja muuttaaksesi käyttöoikeuksia.

Määrittymiset

Määrittymisnäyttö näyttää seuraavat sovellusasetukset:

- **Sovelluksen nimi**
- **Sovelluksen versio**
- **DRAGEN Versio**
- **RTA-versio**
- **Julkaisupäivämäärä**
- **Organisaatio**
- **Laitteen tunniste**
- **Tuotannon tunniste**
- **Library Prep Kits** (Kirjaston valmistelupakkaukset) – Näyttää kirjaston oletusarvoisen valmistelupakkauksen kyseistä sovellusta varten. Tätä asetusta ei voi muuttaa.
- **Index Adapter Kits** (Indeksisovitinsarjat) – Näyttää käytettävissä olevat indeksisovitinsarjat.
- **Indeksiluvut**

- **Lukutyyppi**
- **Indeksin pituudet**
- **Read Lengths** (Lukupituudet) – Lukupituudet asetetaan oletusarvoisesti, kun indeksijoukko valitaan. Tätä asetusta ei voi muuttaa.

Luvat

Nimetyllä järjestelmänvalvojalla on käyttöoikeudet ja hän voi käyttää Käyttöoikeudet-näytön valintaruutuja hallitakseen käyttöoikeuksia TruSight Whole Genome Analysis Application.

Lisätietoja luvista ja käyttäjien hallinnasta on NovaSeq 6000Dx -tuotedokumentaatio (asiakirja nro 200010105):n osassa Järjestelmän määrittäminen.

Ajon luominen

Luo uusia ajoja IVD-tilassa joko laitteessa tai siirtymällä Illumina Run Manager (IRM):een käyttämällä selainta tai verkkoon liitettyä tietokonetta. Voit päästä laitteeseen etätoimintona käyttämällä Illuminan edustajan antamaa osoitetta ja käyttäjätietoa. Katso lisätietoja kohdasta NovaSeq 6000Dx - tuotedokumentaatio (asiakirja nro 200010105).

Create Run (Ajon luominen) on suositeltu ajosuunnittelumenetelmä. Näytelomakkeen tuontia ei suositella. Näytearkin tulostustiedostot ajo- ja analyysikansioissa eivät sovellus tuotavaksi ajosuunnittelun aikana.

Luo ajoja

1. Valitse Runs (Ajot) -näytöstä **Create Run** (Luo ajo).
2. Valitse TruSight Whole Genome Analysis Application 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. Kirjaston valmistelusarja on oletusarvoisesti TruSight Whole Genome, eikä sitä voi muuttaa.
5. Valitse haluamasi TruSight Whole Genome:n indeksisarja avattavasta **Index Adapter Kit** (Indeksisovitinsarja)-pudotusvalikosta.
Lukupituus asetetaan oletusarvoisesti, eikä sitä voi muuttaa. (Luku 1 ja 2 käyttävät 151 jaksoa; Indeksit 1 ja 2 käyttävät 10 jaksoa).
6. Anna kirjaston putkitunnus (suositeltu muoto DX1234567-LIB) ja valitse sitten **Next** (Seuraava). Jos tässä vaiheessa ei ole määritetty kirjaston putkitunnusta, suunniteltu ajo on valittava ennen tarvikkeiden lataamista. Jos virheellinen kirjaston putkitunnus syötetään tässä vaiheessa, suunniteltu ajo on korjattava ennen tarvikkeiden lataamista. Katso kohdasta [Suorita revisio sivulla 5](#) -protokolla, kun haluat korjata ajon ollessasi valmis tarvikkeiden lataamiseen.
7. Näytetiedot- ja Näyteasetukset-näytöissä annetaan näytetiedot. Näytetiedot voidaan syöttää manuaalisesti tai tuomalla näytetiedosto. Näytetunnuksen on oltava yksilöllinen kullekin näytteelle, ja se voi sisältää vain aakkosnumeerisia merkkejä, alaviivoja ja viivoja. Älä sisällytä välilyöntejä. Kaivon sijainti viittaa kaivon indeksilevyn muodossa A01-H04. Indeksisekvenssin tiedot täytetään automaattisesti, kun indeksilevyn kaivon sijainti syötetään. Sukupuoli on annettava muodossa Mies, Nainen tai Tuntematon. Kirjastolevyn tunnus ja kirjaston kaivon tunnus (esim. muoto A01) ovat pakollisia kenttiä.
 - Jos haluat syöttää näytetiedot manuaalisesti, lisää rivit (yhteensä 6 S2- tai 16 S4-virtaussolulle) ja syötä tarvittavat tiedot Sample ID (näytetunniste) ja Well Position (kuopan sijainti) -kenttiin. Tietoja voidaan myös kopioida ja liittää Excelistä. Valitse **Next** (Seuraava). Syötä Näyteasetukset-näytössä Kirjastolevyn tunnus, Kirjaston kaivon tunnus ja sukupuoli. Valitse **Next** (Seuraava).

- Tuo näytedatatieosto valitsemalla **Import Samples** (Tuo näytteet) ja lataamalla näytedatatieosto. Tiedot täytetään riveille automaattisesti. Tällä näytöllä on ladattavissa malli (*.csv). Valitse **Next** (Seuraava). Näyteasetukset-näytössä tiedot täytetään automaattisesti riveille tuodusta näytetiedostosta. Valitse **Next** (Seuraava).
- 8. Syötä Analyysiasetukset-näyttöön erän ja ajon suunnittelun aikana tallennettu Batch Name (erän nimi).
- 9. [Valinnainen] Valitse virtaussolutyyppeiksi, S2 tai S4.
- 10. Vahvista tai poista valinta valintaruudusta, jos haluat luoda ORA-pakatut FASTQ:t, ja valitse sitten **Next** (Seuraava).

HUOMAUTUS TruSight Whole Genome Analysis Application luo ORA-puristetut FASTQ-arvot oletusarvoisesti. Tämän asetuksen muuttaminen suurentaa lopullisen tietojen tuoton kokoa.

- 11. Tarkista annetut tiedot Ajon tarkastelu -näytössä. Jos muutoksia ei tarvita, valitse **Save** (Tallenna). Jos muutoksia tarvitaan, palaa asianmukaiseen näyttöön valitsemalla **Back** (Takaisin).



HUOMIO

TruSight Whole Genome on validoitu 6:lle näytteelle S2 NovaSeq 6000Dx-virtauskennoa käytettäessä ja 16:lle näytteelle NovaSeq 6000Dx S4-virtauskennoa käytettäessä. Varmista, että valittua virtauskenkokokoonpanoa varten on syötetty oikea määrä näytteitä.

Suorita revisio

Jos muutoksia tarvitaan ajon luomisen jälkeen ja ennen tarvikkeiden lataamista sekvensointia varten, tarkista ajot IVD-tilassa joko laitteella tai käyttämällä Illumina Run Manager (IRM) -selainta verkossa olevalla tietokoneella.

1. Valitse **Runs** (Ajot).
2. Valitse Run (Ajon) nimi Planned Runs (Suunnitellut ajot) -välilehdeltä.
3. Valitse **Edit** (Muokkaa).
4. Päivitä ajo- tai näytetiedot tarpeen mukaan. Syötä tai korjaa esimerkiksi kirjaston putkitunnus vastaamaan työkulkua, jota käytettiin työkulun loppuun suorittamisessa.
5. Suorita tarkastelu valitsemalla **Next** (Seuraava) ylös.
6. Valitse **Save** (Tallenna).
7. Valitse **Exit** (Poistu).

Palaa sekvensointiin IVD-tilassa toistaaksesi tarvikkeiden lataamisen. Ajon pitäisi nyt näkyä automaattisesti korostettuna.

Jos päivität kirjaston putken tunnuksen tarvikkeita ladattaessa, palaa Ohjausohjelmiston Suorita valinta -kohtaan ja valitse siihen liittyvän sarakkeen A tai B **Refresh** (Päivitä). Ajon pitäisi nyt olla automaattisesti korostettu. Jos näin ei ole, toista tarvikkeiden lataus valitsemalla **Back** (Takaisin).

Analyysin asettaminen uudelleen jonoon

Katso TruSight Whole Genome Package Insert (asiakirjanro 200050132):n Vianetsintä-osiosta, minkä tyyppinen peräkkäinen analyysi on sopivin.

Analyysin asettaminen uudelleen jonoon ilman muutoksia

1. Valitse Completed Run (suoritetun ajon) nimi tarkastellaksesi Run Details (ajon tietoja).
2. Valitse **Requeue Analysis** (Aseta analyysi uudelleen jonoon).
3. Valitse **Requeue Analysis with no changes** (Aseta analyysi uudelleen jonoon ilman muutoksia).
4. Anna tiedot Reanalysis Reason (Uudelleenanalyysin syy)-kenttään.
5. Valitse **Requeue Analysis** (Aseta analyysi uudelleen jonoon).
6. Poistu sivulta ja siirry Aktiiviset ajot -välilehteen vahvistaaksesi, että analyysin asettaminen jonoon on käynnissä.

Analyysin asettaminen uudelleen muutoksilla

1. Valitse Completed Run (suoritetun ajon) nimi tarkastellaksesi Run Details (ajon tietoja).
2. Valitse **Requeue Analysis** (Aseta analyysi uudelleen jonoon).
3. Valitse **Edit run settings** (Muokkaa ajon asetuksia) ja **Requeue Analysis** (Aseta analyysi uudelleen jonoon).
4. Anna tiedot Reanalysis Reason (Uudelleenanalyysin syy)-kenttään.
5. Valitse **Requeue Analysis** (Aseta analyysi uudelleen jonoon).
6. Vahvista tai päivitä ajoasetukset ja valitse sitten **Next** (Seuraava).
7. Korjaa näytetiedot tarpeen mukaan päivittämällä kentät manuaalisesti tai luo nykyisillä tiedoilla varustettu `sampledata.csv`-tiedosto valitsemalla **Download Template** (Lataa malli). Korjaa tiedot ja poista olemassa olevat rivit Näytetiedot-välilehdellä ennen kuin käytät Tuo näytteet -toimintoa korjattujen näytetietojen täyttämiseen.
8. Tarkastele ajotietoja Tarkastele-näytöllä ja käynnistä uudelleen analysointi valitsemalla **Save** (Tallenna).
9. Valitse **Exit** (Poistu) ja siirry Aktiiviset ajot -sivulle vahvistaaksesi, että jonoon uudelleen asetus käynnissä.

Alkuperäisen ajotietokansion on oltava kohdassa Ajon tiedot määritellyssä ulkoisessa tallennussijainnissa, jotta uudelleenanalysointi onnistuisi. Jos uudelleenanalyysi epäonnistuu, varmista, että ajoa ei ole siirretty tai poistettu.

Ajon ja tulosten tarkasteleminen

1. Valitse IVD-tilan Illumina Run Manager päänäytöstä **Runs** (Ajoj).
2. Valitse Completed Runs (Suoritetut ajot) -välilehdeltä Run (Ajon) nimi.
Tämä välilehti näyttää myös ajot, jotka ovat päättäneet sekvensoinnin, tiedonsiirron tai analyysin epäonnistumisen vuoksi. Aktiiviset ajot ja niiden tila näkyvät Aktiiviset ajot -välilehdessä. Katso lisätietoja kohdasta NovaSeq 6000Dx -tuotedokumentaatio (asiakirja nro 200010105).
3. Valitse Run (Ajon) nimi Completed Runs (Suoritetut ajot) -välilehdeltä, niin näet Run Details and Results (Ajon tiedot ja tulokset) sisältävän Analyysin tulostiedoston polun.
Jos ajo on epäonnistunut, tarkista kunkin vaiheen tila ja katso TruSight Whole Genome Package Insert (asiakirjanro 200050132):n kohta Vianetsintä
4. Siirry paikallisen aseman analyysikansioon ja avaa yhdistetty raportti tarkastellaksesi kunkin laadunvalvontavaiheen HYVÄKSYTTY/HYLÄTTY-tulosta seuraavasti:
 - Katso sekvensointiajon laadunvalvontaa koskevat tiedot kohdasta Laadunvalvonnan tulosten yhteenveto
 - Katso FASTQ QC -kontrollia koskevat tiedot kunkin ajon näytteen osalta kohdasta Yhteenveto FASTQ QC -tulos
 - Kirjaston laadunvarmistus jokaiselle ajon näytteelle on esitetty kohdassa Yhteenveto näytekirjaston laadunvalvonnan tulosJos HYLÄTTY-tulos havaitaan, huomioi laadunvalvontavaihe ja katso TruSight Whole Genome Package Insert (asiakirjanro 200050132):n osaa Vianetsintä.

Tuotostiedoston yhteenveto

TruSight Whole Genome Analysis Application tallentaa seuraavat päätulostetiedostot. Katso päätulostiedostojen sijainti alla olevista tiedostotieto-osioista.

Ajot ja näytteet, jotka eivät läpäise validiteettikriteerejä, eivät tuota CRAM-, ROH- tai *genome.vcf-tiedostoja.

Tuotostiedosto	Kuvaus
Yhdistetty raportti (*.csv)	Sisältää laatumetriikat, joita käytetään ajon validointiin (mukaan lukien kokonaissaanto ja Q30), näytteen validiteettimetriikat (mukaan lukien FASTQ-saanto), kirjaston laadunvalvontametriikat ja vain tiedoksi (FIO) -metriikat kaikille ajon näytteille.
Esimerkkiraportti (*.csv)	Sisältää tulokset sekvensoinnin laadunvalvonnasta, FASTQ:sta ja näytekirjaston laadunvalvonnasta. Raportti sisältää myös ploidisen konkordanssin ja FIO-metriikat yksittäiselle näytteelle sekä siihen liittyvän sekvensointiajon.
Pieni variantti ja mSNV VCF (*.annotated.hard-filtered-gvcf.gz)	Sisältää varianttitunnistustietoja pienille varianteille (SNV, indel-variantit) ja mitokondrio-SNV:ille.
CNV VCF (*.cnv.vcf.gz)	Sisältää varianttitunnistustietoja kopionumeroiden vahvistuksia ja tappioita varten.
Toistaa VCF:n (*.repeats.vcf.gz)	Sisältää varianttitunnistustietoja STR-laajennuksista ja SMN1:stä.
ROH BED (*.roh.bed)	Sisältää homotsygositeettialueiden tietoja.
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.
Kohdistus-CRAM (*.cram)	Sisältää tietyn näytteen kohdistettuja readeja.

Näyteraportin tiedot

Yhdistetty raportti <<RunID>>_Consolidated_Report.csv sijaitsee

TruSightWholeGenomeAnalysis_x.x.x_run-complete -hakemistossa ja sisältää tietoja laatumittareista, joita käytetään näytteiden siirtämiseen tai hylkäämiseen eri analyysivaiheissa.

Yksittäiset näyteraportit <<Sample_ID>>_Sample_Report.csv löytyvät <Sample_ID>-kansioista TruSightWholeGenomeAnalysis_x.x.x_run-complete -hakemistosta.

Raportin otsikoissa on seuraavat tiedot ajosta: sovelluksen versio, erän nimi, kirjaston yhdistelmäputken tunnus, sekvensointiajon nimi, sekvensointiajon tunnus ja virtaussolutyyppi. Seuraavissa taulukoissa kuvataan konsolidoidun raportin sisältämät tiedot. Yksittäinen näyteraportti sisältää samat tiedot lukuun ottamatta Demultiplex-metriikoita.

Taulukko 1 Sekvensoinnin laadunvalvonnan mittarit

Mittarit	Tekniset tiedot	Kuvaus
Ei-indeksoitu kokonaistuotos (Gt)	Ei sovelleta	Ei määrittystä, koska alhaisemmat tuottoajat voivat johtaa näytekirjastojen läpäisemiseen. S4-virtauskennon odotetaan olevan $\geq 3\ 000$ Gbp ja S2-virtauskennon $\geq 1\ 000$ Gt.
Yhteensä % \geq Q30	≥ 85	Emästen laadun mittaaminen ajotasolla. Vähimmäismääritys on asetettu, koska liian alhaiset %Q30-ajot eivät läpäise Q30-emäksiä näytekirjaston laadunvalvonnassa.
Laadunvalvonnan tulosten yhteenveto	HYVÄKSYTTY tai HYLÄTTY	Katso laadunvalvontavirheen sekvensointiohjeet Vianetsintä-osiosta kohdasta TruSight Whole Genome Package Insert (asiakirjanro 200050132).

Taulukko 2 Demultipleksoiva metriikka

Mittarit	Tekniset tiedot	Kuvaus
Prosenttilukemat tunnistettu	Ei sovelleta	Ajon aikana hyväksytyn suodattimen kokonaisuus, joka määritettiin näytteille monipuolisuuden purkamisen aikana.
Prosentti CV	Ei sovelleta	Mittaa ajon jokaisen indeksiparin lukemien tasaisuuden. Odotetaan $< 25\ %$:n ajoille, joissa ei ole FASTQ laadunvalvonta tulosten epäonnistumisia.

Taulukko 3 FASTQ QC -mittarit

Mittarit	Tekniset tiedot	Kuvaus
Saanto näytettä kohden (bps)	$\geq 90\ 000\ 000\ 000$	Minimi on asetettu vastaamaan $\sim 26x$ keskimääräistä autosomaalista peittoa, jotta näytekirjastot voidaan luokitella siten, että ne eivät läpäise laadunvalvontaa analyysiajan lyhentämiseksi.
Yhteenveto FASTQ laadunvalvonta tuloksesta	HYVÄKSYTTY tai HYLÄTTY	Katso FASTQ laadunvalvontavirheen sekvensointiohjeet Vianetsintä-osiosta kohdasta TruSight Whole Genome Package Insert (asiakirjanro 200050132).

Taulukko 4 Näytekirjaston laadunvalvontamittarit

Mittarit	Tekniset tiedot	Kuvaus
Keskimääräinen autosomaalinen kattavuus	≥ 35	Autosomien keskimääräinen kattavuus. Minimimääritys on asetettu analyttisen suorituskyvyn varmistamiseksi.
Prosenttiosuus autosomeista, joiden kattavuus on yli 20X	$\geq 93,94$	Kattavuuden yhdenmukaisuuden mittausta, joka havaitsee ongelmat, jotka eivät välttämättä liity GC-poikkeamaan. Minimimääritys on asetettu analyttisen suorituskyvyn varmistamiseksi.
Normalisoitu kattavuus 60–79 %:n GC-astioissa	$0,82 \leq x \leq 1,13$	Kattavuuden yhdenmukaisuuden mittausta, joka havaitsee GC-poikkeaman, erityisesti kattavuuden menetyksen genomien alueilla, kun GC-prosentti on suurempi ja AT-peruskoostumus pienempi. Minimimääritys on asetettu analyttisen suorituskyvyn varmistamiseksi.
Normalisoitu kattavuus 20–39 %:n GC-astioissa	$0,97 \leq x \leq 1,06$	Kattavuuden yhdenmukaisuuden mittausta, joka havaitsee GC-poikkeaman, erityisesti kattavuuden menetyksen genomien alueilla, kun GC-prosentti on pienempi ja AT-peruskoostumus suurempi. Minimimääritys on asetettu analyttisen suorituskyvyn varmistamiseksi.
Keskimääräinen mitokondriokattavuus	≥ 500	Mitokondriokromosomien kattavuus. Minimimääritys on asetettu varmistamaan mitokondriaalisen SNV:n havaitsemisraja.
Q30-emästen prosenttiosuus	≥ 85	Emästen laadun mittaaminen. Minimimääritys on asetettu analyttisen suorituskyvyn varmistamiseksi.
Arvioitu näytteen saastuminen	$\leq 0,005$	Havaitsee muiden näytteiden saastuttavat lukemat. Maksimimääritys on asetettu varmistamaan mitokondriaalisen SNV:n havaitsemisraja (varianttityyppi, jolla on suurin herkkyys kontaminaatiolle).
Esimerkkikirjaston laadunvalvonnan yhteenvetotulos	HYVÄKSYTTY tai HYLÄTTY	Katso näytekirjaston laadunvalvonnan epäonnistumista Vianetsintä-osiosta kohdasta TruSight Whole Genome Package Insert (asiakirjanro 200050132).

Taulukko 5 Laadunvalvonnan Ploidy-mittarit

Mittarit	Tekniset tiedot	Kuvaus
Annettu sukupuolikromosomiploidia	Ei sovelleta	Käyttäjän antama sukupuoli ajon luomisen aikana (nainen, mies, tuntematon).
Ploidien arviointi	Ei sovelleta	Sukupuoliploidian arvioija DRAGEN.
Yhteenveto Ploidyn tuloksesta	YHTEENSOPIVA, ERILAINEN tai ND	YHTEENSOPIVA osoittaa, että tarjotun ja arvioidun sukupuoliploidian välillä on yhtäpitävyys. ND tarkoittaa sukupuolta, joka on annettu tuntemattomana, tai muuta arviota kuin XX tai XY. ERILAISTEN tulosten osalta joko väärä sukupuoli syötettiin ajon luomisen aikana tai näytteenvaihto on saattanut tapahtua. Katso osaa Vianetsintä kohdasta TruSight Whole Genome Package Insert (asiakirjanro 200050132).

Taulukko 6 Vain tiedoksi -mittarit

Mittarit	Kuvaus
Lisää pituuden mediaani	Tavoite on 450 bp, mutta odottaa vaihtelua sekvensoimalla ajo ja käyttäjä. Alue noin 360–550 bp on hyväksyttävä. Johdonmukainen käyttö tämän vaihteluvälin ulkopuolella voi johtaa suurempaan näytteenottovirheen esiintymistiheyteen.
Prosentteina kartoitetut lukemat	Viitegenomiin kohdistettujen sekvenssifragmenttien prosenttimäärä. Voidaan vähentää vastauksena ei-ihmisen gDNA-kontaminaatioon, näytteen huonoon laatuun tai liian pieneen sisäkkeen pituuteen, mikä johtaa kartoitusongelmiin.
Prosenttia lukemasta täydentävin linjauksin	Prosenttiosuus lukemista, joissa kartoitus jakautuu viitegenomin eri sijainteihin.
Prosentuaalinen kaksoiskappale, merkityt lukemat	Odotettavissa < 20 %. Voidaan nostaa, jos kirjaston valmistelun saanto on alhainen tai poolataan vaadittua volyymia pienemmäksi tai vastauksena sekvensointiin liittyviin ongelmiin.
Pehmeät, leikatut pohjat prosentteina Luku 1	Hyödyllinen keskimääräisen autosomaalisen kattavuuden epäonnistumisen perussyyn diagnosoinnissa.

Mittarit	Kuvaus
Pehmeät, leikatut pohjat prosentteina Luku 2	Hyödyllinen keskimääräisen autosomaalisen kattavuuden epäonnistumisen perussyyn diagnosoinnissa.
Rajattujen emästen prosenttiosuus Luku 1	Hyödyllinen keskimääräisen autosomaalisen kattavuuden epäonnistumisen perussyyn diagnosoinnissa.
Rajattujen emästen prosenttiosuus Luku 2	Hyödyllinen keskimääräisen autosomaalisen kattavuuden epäonnistumisen perussyyn diagnosoinnissa.

Variantin kutsutiedoston tiedot

VCF-tiedostot

Variant call format (*.vcf) -tiedostot sisältävät tietoja varianteista, joita on löydetty tietyissä paikoissa referenssigenomissa, ja ne löytyvät <Sample_ID>/Analysis-hakemistosta.

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

Kaikissa VCF-tiedostoissa on otsikko, jossa on tulossarakkeiden kuvaukset ja varianttikutsutiedot sarakkeissa, jotka on merkitty merkinnöillä CHROM (KROMOS.), POS (ASEMA), ID (TUNNUS), REF (VIITE), ALT (VAIHTOEHT.), QUAL (LAATU), FILTER (SUODATIN), INFO (TIEDOT), FORMAT (MUOTO), SAMPLE (NÄYTE). Sarakkeiden arvojen määritelmät voivat vaihdella varianttien tunnistuksien välillä.

Pieni variantti ja mSNV VCF

Tulos tallennetaan <Sample_ID>.annotated.hard-filtered.gvcf.gz-tiedostoon <Sample_ID>/Analysis-hakemistoon.

Genominen VCF (gVCF) -tiedosto sisältää tietoja varianteista ja asennoista, jotka on määritetty homotsygoottisiksi viitegenomiin nähden. Homotsygoottisilla alueilla gVCF-tiedosto sisältää tilastoja, jotka osoittavat, kuinka hyvin lukemat tukevat varianttien tai vaihtoehtoisten alleelien puuttumista. gVCF-tiedosto sisältää keinotekoisien <NON_REF>-alleelin. Lukemille, jotka eivät tue referenssiä tai mitään varianteja, määritetään <NON_REF>-alleeli. DRAGEN käyttää näitä lukemia määrittämään, voidaanko asemaa tunnistaa homotsygoottiseksi viitteeksi sen sijaan, että se pysyisi tunnistamattomana. Tuloksena saatava pistemäärä edustaa Phred-asteikon mukaista luottamustasoa homotsygoottisessa viitetunnistuksessa. Sukulinjatilassa pisteytys on FORMAT (MUOTO)/GQ.

DRAGEN tarjoaa VCF:n jälkeisen varianttisuodatuksen VCF-tietueissa olevien annotaatioiden perusteella. Varianttien kova suodatus on kuvattu alla. Koska DRAGEN:n algoritmit sisältävät korreloituvien virheiden hypoteesin variantin tunnistajan ytimeistä, putkisto on kuitenkin parantanut kykyä erottaa todelliset variantit melusta, ja siksi riippuvuus VCF:n jälkisuodatuksesta vähenee merkittävästi.

TruSight Whole Genome Analysis Application antaa merkinnän luotettavuuspisteistä ja -tasosta pienille varianteille, joita voidaan käyttää suorituskyvyn parantamiseen. Luottamustason merkintä ei ole laatusuodatin, joten se ei suoraan vaikuta varianttitunnistusten laatuun. Siksi on mahdollista nähdä ohimeneviä varianttitunnistuksia, jotka on kuitenkin merkitty alhaiseksi luottamukseksi.

Taulukko 7 VCF-tiedostojen otsakkeet

Otsake	Kuvaus
CHROM (KROMOS.)	Viitegenomin kromosomi. Kromosomit tulevat näkyviin samassa järjestyksessä kuin referenssi-FASTA-tiedosto.
POS (ASEMA)	Variantin yksittäisen emäksen asema viitekromosomissa. Yhden nukleotidin varianteille (SNV:t) tämä asema on referenssiemäs variantille. Indel-variantteille tämä asema on referenssiemäs, joka välittömästi edeltää varianttia.
ID (TUNNUS)	Aina .
REF (VIITE)	Viitegenotyyppi. Esimerkiksi yksittäisen T:n poisto esitetään viitteenä TT ja vaihtoehtona T. Yksittäinen A - T-nukleotidivariantti esitetään viitteenä A ja vaihtoehtona T.
ALT (VAIHTOEHT.)	Viitereadista poikkeavat alleelit. Esimerkiksi yksittäisen T:n lisäys esitetään viitteenä A ja vaihtoehtona AT. Yksittäinen A - T-nukleotidivariantti esitetään viitteenä A ja vaihtoehtona T.
QUAL (LAATU)	Varianttitunnistimen määrittämä Phred-skaalattu laatupisteytys. Korkeammat pistemäärät viittaavat variantin suurempaan luottamukseen ja virheiden pienempään todennäköisyyteen. Laatupistemäärän Q kohdalla arvioitu virheen todennäköisyys on $10^{-(Q/10)}$. Esimerkiksi Q30-tunnistusten joukon virhemäärä on 0,1 %. Monet varianttitunnistimet määrittävät laatupistemääriä tilastomalliensa perusteella. Ne ovat korkealla tasolla havaitun virhemäärän suhteen.

Taulukko 8 VCF-tiedostomerkinnot

Otsake	Kuvaus
FILTER (SUODATIN)	<ul style="list-style-type: none"> • PASS (HYVÄKSYTTY): Kaikki suodatimet läpäisty. • DRAGENSnpHardQUAL – Aseta, jos tosi:QUAL (LAATU) < 10,41 • DRAGENIndelHardQUAL – Aseta, jos tosi:QUAL (LAATU) < 7,83 • LowDepth – Aseta, jos tosi:DP <= 1 • LowGQ – Määritä, jos tosi:GQ = 0 • PloidyConflict – Genotyypitunnistus varianttitunnistuksesta ei ole yhteneväinen kromosomin ploidian kanssa • base_quality – Sijainti suodatetaan, koska alt-lukujen emäksen mediaanilaatu tässä lokuksessa ei saavuta kynnysarvoa • filtered_reads – Sijainti suodatetaan, koska liian suuri lukujen osuus on poissuodatettu • fragment_length – Sijainti suodatetaan, koska alt-readien fragmentin mediaanipituuden ja ref-readien fragmenttipituuden välinen absoluuttinen erotus tässä lokuksessa ylittää kynnysarvon • low_af – alleelitaajuus ei täytä kynnysarvoa • low_depth – Sijainti suodatetaan, koska read-syvyys on liian pieni • low_frac_info_reads – Sijainti suodatetaan, koska informatiivisten readien osuus alittaa kynnysarvon • low_normal_depth – Sijainti suodatetaan, koska normaalin näytteen read-syvyys on liian pieni • long_indel – Sijainti suodatetaan, koska indel-pituus on liian pitkä • mapping_quality – Sijainti suodatetaan, koska alt-readien kartoituksen mediaanilaatu tässä lokuksessa 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 readia • panel_of_normals – Nähdään vähintään yhdessä näytteessä normaalien vcf-paneelissa • read_position – Sijainti suodatetaan, koska readin ja tämän lokuksen 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 • str_contraction – Sijainti suodatetaan, koska epäillä PCR-virhettä, jossa alt-alleeli on yhden toistoyksikön verran pienempi kuin referenssi • too_few_supporting_reads – Sijainti suodatetaan, koska on liian vähän tukevia readeja kasvainnäytteessä • weak_evidence – Somaattisen variantin pistemäärä ei saavuta kynnysarvoa

Otsake	Kuvaus
INFO (TIEDOT)	<ul style="list-style-type: none"> • 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 read-syvyys (informatiivinen ja ei-informatiivinen); jotkin readit on ehkä suodatettu mapq:n perusteella jne. • END – Välin päättymisasema. • FractionInformativeReads – Informatiivisten readien osuus kokonaisluvuista. • MQ – RMS-kartoituksen laatu. • MQRankSum – Z-pistemäärä Alt vs. Ref -readkartoituksen laatuja Wilcoxonin U-testistä. • ReadPosRankSum – Z-pistemäärä Alt vs. Ref -readkartoituksen aseman systemaattisen virheen Wilcoxonin U-testistä. • SOMATIC – Vähintään yksi variantti tässä asemassa on somaattinen. • IENS – Indel-pituudet kullekin ALT (VAIHTOEHT.)-variantille. • SCORE – Luottamuspisteitys jokaiselle tutkimuskeskuksessa esiintyvälle varianttityypille (varianttityyppi):(luottamuspisteitys). • TIER – Luottamustaso jokaiselle tutkimuskeskuksessa esiintyvälle varianttityypille (varianttityyppi):(luottamustaso).

Otsake	Kuvaus
FORMAT (MUOTO)	<p>FORMAT (MUOTO)-sarakeessa luetellaan kaksoispisteillä erotettuja kenttiä, esimerkiksi GT:GQ.</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 (readit, joissa MQ = 255 tai huonoja pareja, suodatetaan). • F1R2 – Readien määrä F1R2-parin suunnassa, jotka tukevat jokaista alleelia. • F2R1 – Readien määrä F2R1-parin suunnassa, jotka tukevat jokaista alleelia. • GP – Phred-skaalatut jälkitodennäköisyydet genotyypeille, jotka määritellään VCF-tiedoston tiedoissa. • GQ – Genotyypin laatu. • GT – Genotyyppi. • ICNT – INDEL-tietojen lukemien määrä viiteluottamusmallin perusteella. • MB – Näytekohtaisen komponentin tilastotiedot parin systemaattisen virheen havaitsemiseksi. • MIN_DP – GVCF-lauseessa havaittu vähimmäis-DP. • PL – Normalisoidut, Phred-skaalatut todennäköisyydet genotyypeille, jotka määritellään VCF-tiedoston tiedoissa. • PRI – Phred-skaalatut aiemmat todennäköisyydet genotyypeille. • PS – Fysikaalisen tahdistuksen tunnistiedot, 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. • SPL – Normalisoidut, Phred-skaalatut todennäköisyydet SNP:ille viiteluottamusmallin perusteella. • SQ – Somaattinen laatu.
SAMPLE (NÄYTE)	Näytesarakkeessa ilmoitetaan FORMAT (MUOTO) -sarakeessa määritetyt arvot.

Kopiolukuvariantti VCF

Tavoitelaskentavaihe on DRAGEN CNV-putkiston ensimmäinen käsittelyvaihe, joka tuottaa `<Sample_ID>.target.counts.gz`, sitten suoritetaan GC Bias Correction, luoden `*.target.counts.gc-corrected.gz`-tiedoston. Normalisointivaihe tuottaa `*.tn.tsv.gz`-tiedoston. DRAGEN Isäntäohjelmisto luo useita välitiedostoja. `*.seg.called.merged` on lopullinen tunnistustiedosto, joka sisältää vahvistus- ja poistotapahtumat.

Segmenttitiedoston lisäksi tunnistukset DRAGEN lähetetään tavanomaisessa VCF-muodossa. Ulostulo tallennetaan `<Sample_ID>.cnv.vcf.gz`-tiedostoon `<Sample_ID>/Analysis`-hakemistoon.

CNV-tunnistajaa koskevien sarakkeiden määritelmät:

POS (ASEMA)-sarake on variantin aloituskohta. VCF-määrittelyn mukaan, jos jokin ALT (VAIHTOEHT.)-alleeleista on symbolinen alleeli, kuten , pehmusteen pohja on pakollinen ja POS (ASEMA) osoittaa polymorfismia edeltävän pohjan koordinaatin. Kaikki VCF:n koordinaatit ovat 1-pohjaisia.

ID-saraketta käytetään tapahtuman esittämiseen. ID-kenttä koodaa tapahtuman tyyppin ja koordinaatit.

REF-sarake sisältää N:n kaikille CNV-tapahtumille.

ALT-sarake määrittää CNV-tapahtuman tyyppin. Koska VCF sisältää vain CNV-tapahtumia, vain DEL- tai DUP-merkintää käytetään.

QUAL-sarakkeessa on arvioitu laatupisteytys CNV-tapahtumalle, jota käytetään kovassa suodatuksessa.

FILTER-sarakkeessa on PASS (HYVÄKSYTTY), jos CNV-tapahtuma läpäisee kaikki suodattimet. Muussa tapauksessa sarakkeessa on viallisen suodattimen nimi.

INFO-sarake sisältää tapahtumaa edustavia tietoja. REFLLEN-merkintä osoittaa tapahtuman pituuden.

SVTYPE-merkintä on aina CNV. END-merkintä osoittaa tapahtuman lopetussijainnin.

FORMAT (MUOTO)-kentät on kuvattu otsikossa.

- GT – Genotyyppi
- SM – Segmentin keskiarvon lineaarisen kopion suhde
- CN – Arvioitu kopionumero
- BC – Alueen palkkien määrä
- PE – Virheellisesti paritettujen loppulukemien määrä katkoskohtien alussa ja lopussa

Toistaa VCF:n

ExpansionHunter suorittaa sekvenssikuvaajaan perustuvan lukemien uudelleenkohdistuksen kunkin kohteen toiston sisäpuolelta ja ympäriltä. ExpansionHunter sitten genotyypittää toiston pituuden kussakin alleelissa näiden kaavioiden kohdistusten perusteella.

Lisätietoja ja analyseja on saatavilla seuraavista ExpansionHunter-julkaisuista:

- Dolzhenko et al., *Detection of long repeat expansions from PCR-free whole-genome sequence data* 2017
- Dolzhenko et al., *ExpansionHunter: A sequence-graph based tool to analyze variation in short tandem repeat regions* 2019

TruSight Whole Genome Analysis Application STR-varianttikuvaus sisältää tietoja tautia aiheuttavista toistoista, jotka sijaitsevat seuraavissa geeneissä: AFF2, AR, ATN1, ATXN1, ATXN10, ATXN2, ATXN3, ATXN7, ATXN8OS, C9ORF72, CACNA1A, CBL, CNBP, CSTB, DIP2B, DMPK, FMR1, FXN, GLS, HTT, JPH3, NIPA1, NOP56, NOTCH2NL, PABPN1, PHOX2B, PPP2R2B ja TBP.

Toistuvan genotyypityksen tulokset tulostetaan erillisenä VCF-tiedostona, joka antaa kunkin alleelin pituuden kussakin toistettavan määrityksen luettelotiedostossa määritellyssä tunnistettavassa toistossa. Tiedostonimi on <Sample_ID>.repeats.gz, ja se löytyy <Sample_ID>/Analysis-hakemistosta.

Jotkin sarakkeet koskevat toistuvan laajennuksen tunnistajaa:

Taulukko 9 VCF-ydinkentät

Kenttä	Kuvaus
CHROM (KROMOS.)	Kromosomin tunniste
POS (ASEMA)	Ensimmäisen emäksen sijainti ennen toistoaluetta viitteessä
ID (TUNNUS)	Aina .
REF (VIITE)	Referenssipohja aseman sijainnissa
ALT (VAIHTOEHT.)	Luettelo toistuvista alleeleista muodossa <STRn>. N on toistoyksiköiden määrä.
QUAL (LAATU)	Aina .
FILTER (SUODATIN)	LowDepth-suodatinta käytetään, kun lokuksen kokonaissyvyys on alle 10-kertainen tai kun yhden tai molempien murtojen lukemien määrä on alle 5.

Taulukko 10 Muut INFO (TIEFOT) -kentät

Kenttä	Kuvaus
END (LOPETUS)	Viimeisen emäksen sijainti ennen toistoaluetta viitteessä
REF (VIITE)	Viitekopion numero
REPID	Varianttitunnus varianttiluettelosta
RL	Viitepituus muodossa bp
RU	Toista yksikkö viitesuunnassa
VARID	Varianttitunnus varianttiluettelosta

Taulukko 11 GENOTYPE (GENOTYYPPI) (näytettä kohti) -kentät

Kenttä	Kuvaus
AD	Ref- ja alt-alleelien alleelisyvydet luetellussa järjestyksessä
ADFL	Alleelin kanssa yhteensopivien rinnakkaislukujen lukumäärä
ADIR	Alleelin kanssa yhteensopivien toistuvien lukemien lukumäärä
ADSP	Alleelin kanssa yhteensopivien lukukertojen lukumäärä
DST	variantin edustaman testin tulokset (+ havaittu, - havaitsematon, ? määrittämätön)

Kenttä	Kuvaus
GT	Genotyyppi
LC	Lokuksen kattavuus
REPCI	Luottamusväli REPCN:lle
REPCN	Alleelin kattamien toistoyksiköiden määrä
RPL	Todennäköisyysuhde viitealleelin esiintymiselle
SO	Alleelia tukevien lukemien tyyppi. Arvot voivat olla SPANNING, FLANKING tai INREPEAT. Nämä arvot osoittavat, kattavatko lukemat toistoa, sivuavatko ne sitä vai sisältyvätkö ne kokonaan toistoon.

<Sample_ID>.repeats.vcf.gz-tiedosto sisältää SMN-tulosten sekä kaikki kohdennetut toistot. SMN-ulostulo esitetään yksittäisenä SNV-tunnistuksena jatkosta vaikuttavassa asemassa SMN1:ssä (NM_000344.3:c.840C/T), jossa spinaalista lihasatrofiaa (SMA) esiintyy seuraavissa mukautetuissa kentissä.

Taulukko 12 SMA-tulokset toistoissa.vcf-lähtötiedosto

Kenttä	Kuvaus
VARID	SMN merkitsee SMN-tunnistuksen.
GT	Genotyypin tunnistus tässä asennossa käyttäen normaalia (diploidista) genotyyppimallia.
DST	SMA-tilan tunnistus: + osoittaa havaittua - osoittaa havaitsematta jäänyttä ? osoittaa määrittelemätöntä
AD	C- ja T-alleelia tukevat lukumäärät yhteensä.
RPL	Log10-todennäköisyysuhde niiden mallien välillä, joihin tämä ei vaikuta. Positiiviset pisteet osoittavat, että malli, johon ei ole vaikutusta, on todennäköisempi.

ROH BED

Homotsygositeettialueet (ROH) havaitaan osana pienen variantin tunnistajaa. Tunnistaja havaitsee ja tuottaa homotsygositeettiajot ihmisen autosomaalisten kromosomien kokonaisista genomitunnistuksista. Sukupuolikromosomit jätetään huomiotta, ellei näytesukupuolikaryotyyppi ole XX, Ploidy Estimator -arvioijan määrittelyn mukaan. ROH-lähtö mahdollistaa alavirran työkalujen seulomisen ja ennustamisen proband-kohteen vanhempien välillä.

Alue määritetään kromosomin peräkkäisiksi varianttintunnistuksiksi, joissa ei ole suurta väliä näiden varianttien välillä. Toisin sanoen kromosomi tai suuret aukot rikkovat alueet ilman SNV-kutsuja. Raon kooksi on asetettu 3 Mbasea.

ROH-tunnistaja tuottaa ROH-tulostiedoston nimeltä <Sample_ID>.roh.bed <Sample_ID>/Analysis-hakemistossa. Jokainen rivi edustaa yhtä homotsygositeetin aluetta. bed-tiedosto sisältää seuraavat sarakkeet:

```
Chromosome Start End Score #Homozygous #Heterozygous
```

Jossa

- Pisteytys on homotsygoottisten ja heterotsygoottisten varianttien lukumäärän funktio, jossa kukin homotsygoottinen variantti nostaa pistemäärää 0,025:llä ja jokainen heterotsygoottinen variantti vähentää pistemäärää 0,975:llä.
- Aloitus- ja lopetusasennot ovat 0-pohjaisia, puoliavoin väli.
- #Homotsygoottinen on alueen homotsygoottisten varianttien määrä.
- #Heterotsygoottinen on alueella olevien heterotsygoottisten varianttien määrä.

Tunnistaja tuottaa myös metriikkatiedoston nimeltä <Sample_ID>.roh_metrics.csv, joka luettelee suuren ROH:n määrän ja SNP-prosenttiosuuden suuressa ROH:ssa (> 3 Mt).

Ploidyn arvioinnin metriikat

Ploidy Estimator toimii oletusarvoisesti. Ploidy Estimator laskee sekvensointisyvyyden ihmisgenomin jokaisen autosoomin ja allosomin osalta kartoitus-/kohdistimen lukemien avulla. Näytteen sukupuolikaryotyyppi arvioidaan sitten käyttämällä seksuaalikromosomipeittojen mediaanin ja autosomaalisen peittoalueen mediaanin suhdetta. XX tai XY ja YHTEENSOPIVA, ERILAINEN tai ND (ei määritetty). verrattuna annettuihin näytetietoihin raportoidaan konsolidoidussa raportissa. Yksityiskohtaiset tulokset, mukaan lukien jokainen normalisoitu jatkuva mediaanikattavuus, raportoidaan <Sample_ID>.ploidy_estimation_metrics.csv-tiedostossa.

FASTQ-tiedostot

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

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

Ohjelmisto luo yhden FASTQ-tiedoston jokaista näytettä, lukua ja kaistaa varten. Ohjelmisto luo esimerkiksi jokaisesta pariliitetyn suoritusjärjestyksen näytteestä kaksi FASTQ-tiedostoa: yksi lukua 1 ja toinen lukua 2 varten. Näiden FASTQ-näytetiedostojen lisäksi ohjelmisto luo kaksi FASTQ-tiedostoa kaistaa kohti sisältäen kaikki tuntemattomat näytteet. FASTQ-tiedostoja indeksilukua 1 ja indeksilukua 2 varten ei luoda, koska sekvenssi sisältyy kunkin FASTQ-merkinnän otsikkoon. Tiedostonimen muoto muodostetaan näytearkissa määritetyistä kentistä ja käytetään tiedostonimen muotoa <Sample_ID>_S#_L00#_R#_001.fastq.gz

FASTQ-tiedostot tallennetaan <Sample_ID>/Conversion-hakemistoon. Analyysikansion FASTQ-hakemistosta löytyy lokihakemisto, jossa on BCL-to-FASTQ-muuntolokit, ja raporttihakemisto, joka sisältää erilaisia lukemetriikkatiedostoja, ja SampleSheet.csv, jota käytetään FASTQ-muunnokseen. FASTQ-tiedostot, jotka ovat peräisin määrittämättömistä lukemista, löytyvät analyysikansion Undetermined/Conversion hakemistoista.

Näytteen tunnisteiden muoto on seuraava:

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber
Example:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<?????#=#
```

CRAM-tiedostot

Pakatut viitesuunnatut kohdistuskartta- tai CRAM-tiedostot (*.cram) tallennetaan <Sample_ID>/Analysis-hakemistoon ja sisältävät otsikkoja ja kohdistustietueita kohdistuksen aikana käytettyyn genomiseen viitetiedostoon nähden. Viitetiedoston polku on lueteltu <Sample_ID>/Analysis/<Sample_ID>-replay.json-tiedostossa --ht-reference-parametrina, jonka oletusasetuksena on hg38.fa.

CRAM-tiedostot sisältävät otsikko-osion ja kohdistusosion:

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

Kohdistusosioon sisältyvät seuraavat tiedot kutakin readia tai read-paria varten:

- AS: Paired-end-kohdistuksen laatu.
- RG: Readin ryhmä, joka osoittaa tietyn näytteen readien määrän.
- BC: Viivakooditunniste, joka osoittaa readiin liittyvän demultipleksoidun näytteen tunnuksen.
- SM: Yhden pään kohdistuksen laatu.
- XC: Osuman kuvaajan merkkijono.
- XN: Amplikoninimitunniste, joka tallentaa readiin liittyvän amplikonin tunnuksen

Kohdistustietueiden tarkastelussa samtools voidaan käyttää samtools view --reference <path_to_reference_folder>/hg38.fa <Sample_ID>.cram.

Myös indeksitiedosto ja tarkistussummatiedosto luodaan.

Tekninen tuki

Teknisissä kysymyksissä voit ottaa yhteyttä Illuminan tekniseen tukeen.

Verkkosivusto: www.illumina.com

Sähköposti: techsupport@illumina.com

Käyttöturvallisuustiedotteet (SDS) — saatavilla Illuminan verkkosivustolta osoitteesta support.illumina.com/sds.html.

Tuotedokumentaatio — ladattavissa verkkosivustolta support.illumina.com.



Illumina, Inc.
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



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

Rahoittaja Australiassa

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

DIAGNOSTISEEN IN VITRO -KÄYTTÖÖN.

© 2024 Illumina, Inc. Kaikki oikeudet pidätetään.

illumina[®]