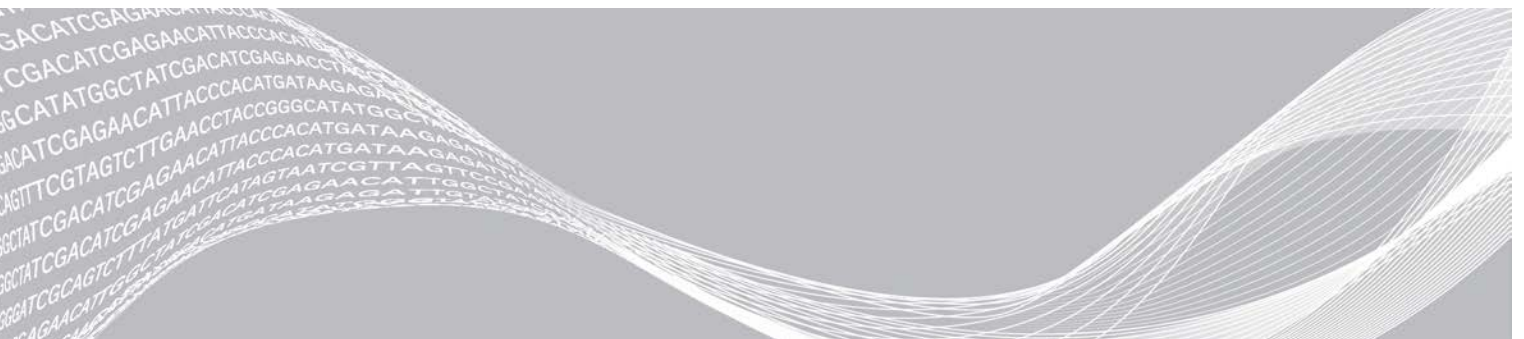


# Local Run Managerin ituratavarianttianalyysimoduulin

työnkulkuopas NextSeq 550Dx:lle

IN VITRO -DIAGNOSTISEEN KÄYTTÖÖN

Yleiskatsaus	3
Ajotietojen syöttäminen	3
Analyysimenetelmät	6
Ajo- ja näytetietojen tarkastelu	7
Analyysiraportti	7
Analyysin tuotostiedostot	9
Versiohistoria	16
Tekninen tuki	17



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).

© 2021 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](http://www.illumina.com/company/legal.html).

## Yleiskatsaus

Local Run Managerin ituratavarianttimoduuli on tarkoitettu käytettäväksi Illuminan mukautetun TruSeq-Custom Amplicon Kit Dx -määrityksen ja NextSeq 550Dx:n kanssa. Ituratavarianttimoduulin kanssa käytettynä määritys on tarkoitettu valmistelemaan kirjastoja, joita käytetään DNA:n sekvensointiin perifeerisistä kokoverinäytteistä.

Analyysimoduulilla arvioidaan lyhyitä alueita, joilla on varianttien amplifioitua DNA:ta tai ampliconeja. Ampliconien fokusoidun sekvensoinnin avulla tiettyjen alueiden suuri kattavuus on mahdollinen. Katso pakkausseloste *Mukautettu TruSeq Custom Amplicon Kit Dx (asiakirjanro 1000000029772)*.

Ituratavarianttianalyysimoduuli edellyttää 300 jakson sekvensointitarvikkeita. Katso lisätietoja *NextSeq 550Dx -suurtehoreagenssisarjan v2* tai *NextSeq 550Dx -suurtehoreagenssisarjan v2.5* pakkausselosteesta.

## Tietoa tästä oppaasta

Tässä oppaassa annetaan ohjeet ajoparametrien määrittämiseksi sekvensointia ja ituratavarianttianalyysimoduulin analysointia varten. Tietoa Local Run Managerin koontinäyttö- ja järjestelmän asetuksista on *NextSeq 550Dx -instrumentin viiteoppaassa (asiakirja nro 1000000009513)*.

## Local Run Managerin tarkastelu

Local Run Manager -liittymää tarkastellaan NextSeq 550Dx -käyttöohjelmistolla (NOS) tai verkkoselaimella. Tuettu verkkoselain on Chromium.



### HUOMAUTUS

Mikäli käytät tukematonta selainta, lataa tuettu selain, kun siihen kehoitetaan Confirm Unsupported Browser (Vahvista tukematon selain) -viestissä. Valitsemalla "**here**" (tästä) voit ladata tuetun Chromium-version.

## Tarkastele laitteen näytöllä

- 1 Jotta voit tarkastella Local Run Manager -liittymää laitteen näytöllä, valitse yksi seuraavista vaihtoehdoista:
  - ▶ Valitse NOS -aloitusnäytöltä **Local Run Manager** (Paikallinen ajonhallinta). Kun olet valmis, voit palata NOSiin napsauttamalla oikean yläkulman X-kuvaketta.
  - ▶ Valitse Minimize NOS (Minimoi NOS) -kuvake, avaa laitteessa oleva Chromium-verkkoselain ja syötä osoiteriville **http://localhost**. Vain järjestelmänvalvojat voivat minimoida NOSin.

## Tarkastele verkostoon kuuluvalta tietokoneelta

- 1 Avaa Chromium-verkkoselain tietokoneella, jolta on pääsy käyttämään samaa verkostoa kuin laitteelta, ja muodosta yhteys instrumentin IP-osoitteen tai instrumentin nimen avulla. Esimerkiksi **http://myinstrument**.

## Ajotietojen syöttäminen

### Parametrien määrittäminen

- 1 Kirjautu sisään Local Run Managerissa.
- 2 Valitse **Create Run** (Luo ajo) ja valitse **Germline Variant** (Ituratavariantti).

- 3 Syötä ajonimi, jossa ajo tunnistetaan sekvensoinnista analyysin kautta. Käytä alfanumeerisia merkkejä, välilyöntejä, alaviivoja tai väliviivoja.
- 4 **[Valinnainen]** Syötä ajon kuvaus ajon tunnistamisen helpottamiseksi. Käytä alfanumeerisia merkkejä, välilyöntejä, alaviivoja tai väliviivoja.
- 5 Valitse näytteiden määrä ja indeksijoukko pudotusvalikosta. Ota valintaa tehdessäsi huomioon seuraavat tiedot.
  - ▶ Pudotusvalikko sisältää näytteiden määrät yhdessä indeksijoukon kanssa. Esimerkiksi 24-joukko 1 viittaa 24 testattavaan näytteeseen indeksijoukon 1 indekseillä.
  - ▶ Indeksijoukon numeroilla viitataan eri i5-indeksijoukkoihin. Sekä joukko 1 että joukko 2 tarjoavat indeksidiversiteettiä. Tarjotaan kaksi indeksijoukkoa avuksi yksittäisen joukon loppuunkulumisen estämiseen.
  - ▶ Valitse se näytemäärä, joka on lähinnä testaamiesi näytteiden määrää. Mikäli näytteiden tarkkaa määrää ei ole luettelossa, valitse numero, joka on lähinnä testaamaasi numeroa mutta sitä pienempi. Jos esimerkiksi haluat testata 18 näytettä, valitse 16 näytettä.
  - ▶ Näytteen kuopat ja indeksiyhdistelmät, jotka vastaavat indeksin diversiteettivaatimuksia, korostetaan vihreällä. Jos valitset muita kuoppa- ja indeksiyhdistelmiä, saat ajoa tallentaessasi ilmoituksen, jos indeksin diversiteettivaatimukset eivät täyty.

## Manifestitiedostojen tuonti ajoa varten

- 1 Varmista, että manifestitiedostot, jotka haluat tuoda, ovat saatavilla käytettävissä olevassa verkkosijainnissa tai USB -asemassa.
- 2 Valitse **Import Manifests** (Tuo manifestitiedostot).
- 3 Navigoi manifestitiedoston kohdalle ja valitse manifestitiedostot, jotka haluat lisätä.



### HUOMAUTUS

Voit asettaa manifestitiedostot kaikkien ajojen saataville ituratavarianttianalyysimoduulin avulla ja lisätä manifestitiedostoja Module Settings (Moduulin asetukset) -ominaisuuden avulla. Tämä ominaisuus edellyttää järjestelmänvalvojan käyttöoikeuksia. Katso lisätietoja *NextSeq 550Dx -instrumentin viiteoppaasta (asiakirjanro 1000000009513)*.


## Näytteiden määrittäminen ajoa varten

Määritä näytteet ajoa varten käyttämällä yhtä seuraavista vaihtoehtoista ja ohjeista.


- ▶ **Enter samples manually** (Syötä näytteet manuaalisesti) — Käytä tyhjää taulukkoa Create Run (Luo ajo) -näytöllä.
- ▶ **Import samples** (Tuo näytteet) — Navigoi ulkoiseen tiedostoon muodossa, jossa käytetään pilkulla erotettuja arvoja (\*.csv). Create Run (Luo ajo) -näytöllä on saatavilla ladattava malline.

Kun olet täyttänyt näytetaulukon, voit viedä näytetiedoston ulkoiseen tiedostoon. Käytä tiedostoa viitteenä, kun valmistelet kirjastoja tai tuot tiedoston muuta ajoa varten.

## Näytteiden syöttäminen manuaalisesti

- 1 Syötä ainutkertainen näytetunnus Sample ID (Näytetunnus) -kenttään.  
Käytä alfanumeerisia merkkejä, välilyöntejä tai alaviivoja.
- 2 **[Valinnainen]** Napsauta positiivisten tai negatiivisten kontrollinäytteiden kohdalla hiiren oikeaa painiketta ja valitse kontrollityyppi.
- 3 **[Valinnainen]** Syötä näytekuvaus Sample Description (Näytekuvaus) -kenttään.  
Käytä alfanumeerisia merkkejä, välilyöntejä tai alaviivoja.  
Näytekuvaukset yhdistetään näytenimeen. Näytekuvaukset korvataan, jos samaa näytenimeä käytetään uudelleen myöhemmässä ajossa.
- 4 Valitse Indeksien 1 sovitin Indeksien 1 (i7) pudotusvalikosta.  
Kun käytät ehdotettuja näytekuoppia, ohjelmisto täyttää automaattisesti i7- ja i5-indeksisovittimet, jotka vastaavat diversiteetti-indeksien vaatimuksia. Mikäli testaamiesi näytteiden tarkka määrä ei ole luettelossa, varmista, että valitaan indeksisovittimet ylimääräisille kuopille. Jos on valittava indeksejä ylimääräisille kuopille tai jos et käytä suositeltuja indeksisovitinyhdistelmiä, varmista ennen indeksien valitsemista, että luet *Emästen tunnistaminen ja indeksin monimuotoisuus sivulla 15*.
- 5 Valitse Indeksien 2 sovitin Indeksien 2 (i5) pudotusvalikosta.
- 6 Valitse manifestitiedosto Manifest (Manifesti) -pudotusvalikosta.
- 7 Valitse vaihtoehto levyasettelun tarkastelemiseksi, tulostamiseksi tai tallentamiseksi viitteenä kirjastojen valmistelua varten:
  - ▶ Voit näyttää levyasettelun valitsemalla  **Print** (Tulosta) -kuvakkeen. Voit tulostaa levyasettelun valitsemalla **Print** (Tulosta).
  - ▶ Valitsemalla **Export** (Vie) voit viedä näytetiedot ulkoiseen tiedostoon.  
Varmista, että manifestitiedot ja näytetiedot ovat oikeat. Virheelliset tiedot saattavat vaikuttaa tuloksiin.
- 8 Valitse **Save Run** (Tallenna ajo).

## Näytteiden tuonti

- 1 Valitse **Import Samples** (Tuo näytteet) ja selaa näytteen tietotiedoston kohtaan. Voit tuoda kahdentyyppisiä tiedostoja.
  - ▶ Valitse **Template** (Malline) Create Run (Luo ajo) -näytöltä, jotta voit luoda uuden levyasettelun. Mallinetiedosto sisältää oikeat sarakeotsikot tuontia varten. Syötä näytetiedot kuhunkin ajon näytteiden sarakeeseen. Poista esimerkkitiedot käyttämättömistä soluista ja tallenna sitten tiedosto.
  - ▶ Käytä näytetietotiedostoa, joka vietiin ituratavarianttimoduulista Export (Vienti) -ominaisuuden avulla.
- 2 Voit näyttää levyasettelun valitsemalla  **Print** (Tulosta) -kuvakkeen.
- 3 Valitse **Print** (Tulosta) levyasettelun tulostamiseksi viitteeksi kirjastojen valmistelua varten.
- 4 **[Valinnainen]** Valitsemalla **Export** (Vie) voit viedä näytetiedot ulkoiseen tiedostoon.  
Varmista, että manifestitiedot ja näytetiedot ovat oikeat. Virheelliset tiedot saattavat vaikuttaa tuloksiin.
- 5 Valitse **Save Run** (Tallenna ajo).

## Ajon muokkaaminen

Katso ennen sekvensointia tapahtuvaa ajon tietojen muokkausta koskevat ohjeet *NextSeq 550Dx -instrumentin viiteoppaasta (asiakirjanro 1000000009513)*.

## Analyysimenetelmät

Ituratavarianttianalyysimoduuli suorittaa seuraavat analyysivaiheet ja kirjoittaa sitten analyysin tuotostiedostot Alignment (Kohdistus) -kansioon.

- ▶ Demultipleksoidaan indeksireadit
- ▶ Luodaan FASTQ-tiedostot
- ▶ Kohdistetaan viitteeseen
- ▶ Tunnistetaan variantit

## Demultipleksointi

Demultipleksoinnissa verrataan kutakin Index Read -sekvenssiä ajoa varten määritettyihin 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 määrätään näytteeseen, kun indeksisekvenssi on tarkka osuma tai kun indeksireadia kohti on enintään yksi yhteensopimattomuus.

## FASTQ-tiedoston luonti

Demultipleksoinnin jälkeen ohjelmistossa luodaan välianalyysitiedostoja FASTQ-muodossa. Se on tekstimuoto, jota käytetään sekvenssien esittämiseen. FASTQ-tiedostot sisältävät readeja kustakin näytteestä ja niihin liittyvistä laatupesteytyksistä. Klusterit, jotka eivät läpäisseet suodatinta, jätetään pois.

Jokainen FASTQ-tiedosto sisältää readeja vain yhdestä näytteestä, ja kyseisen näytteen nimi sisällytetään FASTQ-tiedoston nimeen. FASTQ-tiedostot ovat kohdistuksen pääasiallisia syötetietoja. Näytettä kohti luodaan FASTQ-tiedostoa, neljä Readista 1 ja neljä Readista 2.

## Kohdistus

Kohdistusvaiheen aikana Smith-Waterman-algoritmi kohdistaa kunkin näytteen klusterit manifestitiedostossa määritettyihin amplikonisekvensseihin nähden.

Porrastettu Smith-Waterman-algoritmi suorittaa puolittain globaaleja sekvenssikohdistuksia määrittääkseen samankaltaisia alueita kahden sekvenssin välissä. Koko sekvenssin vertailemisen sijasta Smith-Waterman-algoritmillä verrataan kaikenpituisia segmenttejä.

Jokaista parillisen pään readia arvioidaan suhteessa sen kohdistukseen kyseisen readin asiaankuuluviin sondisekvensseihin nähden.

- ▶ Read 1 arvioidaan loppupään lokusspesifisten oligojen (Downstream Locus-Specific Oligos I. DLSO) käänteiseen täydennykseen nähden.
- ▶ Read 2 arvioidaan alkupään lokusspesifisiin oligoihin (Upstream Locus-Specific Oligos I. ULSO) nähden.
- ▶ Mikäli readin alku vastaa sondisekvenssiä, jossa on enintään kolme eroa (johtavista indeleistä aiheutuvaa yhteensopimattomuutta tai siirtymää), readin koko pituus kohdistetaan kyseisen sekvenssin amplikonikohteeseen nähden.
- ▶ DLSO:n ja ULSO:n sisäisiä indeleitä ei havainnoida määrittämisen kemia huomioon ottaen.

Kohdistukset suodatetaan kohdistustuloksista yhteensopimattomuusasteiden perusteella joko kohdealueelta tai koko amplikonista amplikonin pituuden mukaan. Suodatetut kohdistukset kirjoitetaan kohdistustiedostoihin kohdistamattomina, eikä niitä käytetä varianttien tunnistamiseen.

## Varianttintunnistus

Illuminan kehittämä Pisces-varianttintunnistin tunnistaa variantit, joita on DNA-näytteessä.



Pisces-varianttintunnistimella tunnistetaan SNV:t, MNV:t ja pienet indelit kolmessa vaiheessa:

- ▶ Otetaan huomioon erikseen viitegenomin jokainen asema
- ▶ Lasketaan emäkset tietyssä asemassa sellaisten kohdistettujen readien kohdalla, jotka menevät aseman kanssa päällekkäin
- ▶ Lasketaan varianttitulos, jolla mitataan Poisson-mallia hyödyntävän tunnistuksen laatua. Variantit, joiden laatuasteetta on alle Q20, suljetaan pois.

Mikäli variantti läpäisee kaikki suodattimet, sen kohdalle merkitään PASS (HYVÄKSYTTY) VCF:ssä.

Lisätietoja on osoitteessa [github.com/Illumina/Pisces/wiki](https://github.com/Illumina/Pisces/wiki).

## Ajo- ja näytetietojen tarkastelu

- 1 Napsauta Local Run Manager -koontinäytöltä ajon nimeä.
- 2 Tarkista sekvensointiajon mittarit Run Overview (Ajon yleiskatsaus) -välilehdeltä.
- 3 **[Valinnainen]** Napsauttamalla  **Copy to Clipboard** (Kopioi leikepöydälle) -kuvaketta voit kopioida tuotosajon kansiopolon.
- 4 Napsauttamalla Sequencing Information (Sekvensointitiedot) -välilehteä voit tarkistaa ajoparametrit ja tarvike tiedot.
- 5 Napsauttamalla Samples and Results (Näytteet ja tulokset) -välilehteä voit tarkastella analyysiraportin sijaintia.
  - ▶ Mikäli analyysi toistettiin, laajenna Select Analysis (Valitse analyysi) -pudotusvalikkoo ja valitse asianmukainen analyysi.
- 6 Napsauttamalla  **Copy to Clipboard** (Kopioi leikepöydälle) -kuvaketta voit kopioida Analysis (Analyysi) -kansiopolon.

Lisätietoja Run Overview (Ajon yleiskatsaus)- ja Sequencing Information (Sekvensointitiedot) -välilehdistä ja analyysin asettamisesta uudelleen on *NextSeq 550Dx -instrumentin viiteoppaassa (asiakirjanro 1000000009513)*.

## Analyysiraportti

Analyysitulokset tiivistetään Samples and Results (Näytteet ja tulokset) -välilehdellä ja koottuna raporttina Alignment (Kohdistus) -kansiossa. Raportti kustakin näytteestä on myös saatavilla PDF-tiedostomuodossa kunkin näytteen kohdalla.

## Samples and Results (Näytteet ja tulokset -välilehtitiedot)

- 1 Voit tarkastella näyteraporttia napsauttamalla luettelossa olevaa näytettä.

**Taulukko 1 Ajo- ja näytetiedot**

Sarakkeen otsikko	Kuvaus
Run Status (Ajon tila)	Osoittaa, onko sekvensointiajo hyväksytty vai hylätty.
Total Yield (GB) (Kokonaistuotos (Gt))	Sekvensointiajossa tunnistettujen emästen määrä. Näyttää hyväksyntäkynnyksen sekä hyväksytyt tai hylätyt tilat.
% ≥ Q30	Sekvensointiajon readien prosenttiosuus sekä laatupesteytys, joka on vähintään 30(Q30). Näyttää hyväksyntäkynnyksen sekä hyväksytyt tai hylätyt tilat.
Sample ID (Näytetunnus)	Näytetunnus, joka ilmoitettiin ajoa luotaessa.
Total PF Reads (PF-readit yhteensä)	Suodattimen läpäisevien readien määrä yhteensä.
Read 1 % ≥ Q30	Readin 1 readien prosenttiosuus sekä laatupesteytys, joka on vähintään 30(Q30) näytteen kohdalla.
Read 2 % ≥ Q30	Readin 2 readien prosenttiosuus sekä laatupesteytys, joka on vähintään 30(Q30) näytteen kohdalla.
Autosome Call Rate (Autosomin tunnistusaste)	Genomiasemien määrä autosomien halki (kromosomit 1 - 22), jotka vastaavat ennalta määritetyn luottamusarvon kynnystä jaettuna tutkittujen autosomaalisten genomiasemien kokonaismäärällä. Tunnistusastetta kuvataan näytekohtaisesti ja se ilmoitetaan prosenttiosuutena, joka lasketaan 1 miinuksena (sellaisten autosomaalisten asemien määrä, joihin liittyy keskeneräisiä tunnistuksia, jaettuna sekvensoitujen autosomaalisten asemien kokonaismäärällä).

**Taulukko 2 Näyteraportin tiedot**

Sarakkeen otsikko	Kuvaus
Sample (Näyte)	Näytetunnus, joka ilmoitettiin ajoa luotaessa.
Report Date (Raportin päivämäärä)	Raportin luontipäivämäärä.
Sample Information (Näytetiedot)	Näytetunnus, joka ilmoitettiin ajoa luotaessa, näytteessä suodattimen läpäisseet readit yhteensä, näytteen sellaisten readien prosenttiosuus, joiden laatupestemäärä on vähintään 30(Q30), sekä autosomaalinen tunnistusaste.
Amplicon Summary (Amplikonitiivistelmä)	Sekvensoitujen amplikonialueiden kokonaismäärä sekä sekvensoitujen amplikonien emäsparien kokonaispituus kohdealueilla näytteen sekä manifestitiedoston kohdalla. Manifestitiedostossa määritetään viitegenomi ja kohteena olevat viitealueet, joita käytetään kohdistusvaiheessa.
Read Level Statistics (Read-tasotilastot)	Näytteen numero ja näytteen readien prosenttiosuus, jotka kattavat viitteen kaikki asemat Readin 1 ja Readin 2 kohdalla.
Variants Summary (Varianttiyhteenveto)	Näytteen kohdalla havaittujen SNV-kohteiden, lisäysten ja poistojen määrä, jotka läpäisivät ehdotetut arvot, jotta voidaan määrittää, ovatko laatuoloskset hyväksyttävällä alueella.
Coverage Summary (Kattavuustiivistelmä)	Kohdistettujen emästen kokonaismäärä jaettuna kohteena olevan alueen koolla sekä sellaisten amplikonialueiden prosenttiosuus, joiden kattavuusarvot ovat suurempia kuin alhaisen kattavuuden kynnyksen, amplikonin keskimääräinen kattavuus 0,2 * näytteen kohdalla.
Coverage Plots (Kattavuuskuviot)	Amplikonialuekohtaisessa kattavuuskaaviossa esitetään kattavuus näytteen amplikonialueiden poikki. Alueet, joiden kattavuusarvot ovat kattavuuskynnystä alhaisemmat, on korostettu punaisella. Kaikkien arvojen keskiarvo osoitetaan oranssilla viivalla.
Software Versions (Ohjelmistoversiot)	Ohjelmistoversiot näytteen sekvensoinnin aikana. Sisältää NextSeq 550Dx -käyttöohjelmiston (NOS), Local Run Manager -ohjelmiston, RTA-ohjelmiston ja ituratavarianttimoduuliversion.



## Analyysin tuotostiedostot

Seuraavat analyysin tuotostiedostot luodaan ituratavarianttianalyysimoduulia varten ja niistä saadaan analyysitulokset kohdistukseen ja varianttien tunnistukseen. Analyysin tuotostiedostot sijaitsevat Alignment (Kohdistus) -kansiossa.

Tiedostonimi	Kuvaus
Demultiplexing (*.txt)	Välitiedostot sisältävät demultipleksoinnin tiivistetyt tulokset.
FASTQ (*.fastq.gz)	Välitiedostot, jotka sisältävät laatupisteytykseen perustuvia emästen tunnistuksia. FASTQ-tiedostot ovat kohdistusvaiheen pääasiallisia syötetietoja.
Kohdistustiedostot BAM-muodossa (*.bam)	Sisältää tietyn näytteen kohdistettuja readeja.
Varianttien tunnistustiedostot genomien VCF-muodossa (*.genome.vcf.gz)	Sisältää kunkin aseman genotyypin riippumatta siitä, tunnistetaanko ne varianteiksi vai viitteiksi.
Varianttien tunnistustiedostot VCF-muodossa (*.vcf.gz)	Sisältää kaikki kohdealueella tunnistetut variantit.
AmpliconCoverage_M1.tsv	Sisältää tietoa amplikoni- ja näytekohtaisesta kattavuudesta kunkin toimitetun manifestitiedoston kohdalla. M# viittaa manifestitiedoston numeroon.

## Demultipleksointi-tiedostomuoto

Demultipleksointiprosessissa luetaan kuhunkin klusteriin liitetty hakemistosekvenssi klustereiden alkuperänäytteen määrittämiseksi. Klustereiden ja näytenuumeron välinen kartoitus kirjoitetaan demultipleksointitiedostoon (\*.demux) virtauskyvetin jokaisen ruudun kohdalla.

Demultipleksointitiedoston nimeämismuoto on `s_1_X.demux`, jossa X on ruudun numero.

Demultipleksointitiedostojen alussa on otsikko:

- ▶ Versio (4-tavuinen kokonaisluku), tällä hetkellä 1
- ▶ Klusterimäärä (4-tavuinen kokonaisluku)

Tiedoston muu osa koostuu ruudun kunkin klusterin näytenumeroista.

Kun demultipleksointivaihe on valmis, ohjelmisto luo demultipleksointitiedoston nimeltä `DemultiplexSummaryF1L1.txt`.

- ▶ Tiedostonimessä **F1** edustaa virtauskyvetin numeroa.
- ▶ Tiedostonimessä **L1** edustaa kaistan numeroa.
- ▶ Demultipleksoinnin tuloksena syntyy taulukko, jossa on 1 rivi ruutua kohti ja 1 sarake näytettä kohti, näyte 0 mukaan luettuna.
- ▶ Indeksireadien yleisimmin ilmenevät sekvenssit.

## FASTQ-tiedostomuoto

FASTQ on tekstipohjainen tiedostomuoto, joka sisältää emästen tunnistuksia ja laatuarvoja readia kohti. Jokaisessa tietueessa on 4 riviä:

- ▶ Tunniste
- ▶ Sekvenssi

- ▶ Plus-merkki (+)
- ▶ Phred-laaturipisteytykset ASCII + 33 -koodatussa muodossa

Tunniste muotoillaan seuraavasti:

**@Instrumentti:Ajotunniste:Virtauskyvettitunniste:Kaista:Ruutu:X:Y Read-numero:Suodatinlippu:0:Näytenumero**

Esimerkki:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<?????#=#
```

## BAM-tiedostomuoto

BAM-tiedosto (\*.bam) on kompressoitu binääri-versio SAM-tiedostosta, jota käytetään esittämään kohdistettuja sekvenssejä 128 Mt:uun saakka. SAM- ja BAM-muodot on kuvattu yksityiskohtaisesti tiedostossa [samtools.github.io/hts-specs/SAMv1.pdf](https://samtools.github.io/hts-specs/SAMv1.pdf).

BAM-tiedostoissa käytetään tiedostojen nimeämismuotoa **NäyteNimi\_S#.bam**, jossa # on näytteen numero määritettynä järjestyksessä, jossa näytteet luetellaan ajoa varten.

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:** Parillisen pään kohdistuslaatu
- ▶ **RG:** Read-ryhmä, joka osoittaa tietyn näytteen readien määrän.
- ▶ **BC:** Viivakooditunniste, joka osoittaa readiin liittyvän demultipleksoidun näytteen tunnuksen.
- ▶ **SM:** Yksittäisen pään kohdistuslaatu.
- ▶ **XC:** Osuman kuvausmerkkijono
- ▶ **XN:** Amplikoninimitunniste, joka tallentaa readiin liittyvän amplikonin tunnuksen

BAM-hakemistotiedostoissa (\*.bam.bai) on vastaavan BAM-tiedoston hakemisto.

## VCF-tiedostomuoto

Varianttitunnistusmuoto (VCF) on genomiikkatiedeyhteisön kehittämä yleinen tiedostomuoto.

Se sisältää tietoa viitegenomin tietyistä asemista löytyneistä varianteista. VCF-tiedostoissa on .vcf-päätte

VCF-tiedoston otsikkoon sisältyy VCF-tiedostomuoto-versio ja varianttitunnistinversio, ja siinä luetellaan tiedoston muussa osassa käytetyt merkinnät. VCF-otsikkoon sisältyy myös viitegenomitiedosto ja BAM-tiedosto. Otsikon viimeinen rivi sisältää tietorivien sarakkeiden otsakkeet. Jokainen VCF-tiedoston tietorivi sisältää tietoa yhdestä variantista.

## VCF-tiedostojen otsakkeet

Otsake	Kuvaus
CHROM	Viitegenomin kromosomi. Kromosomit tulevat näkyviin samassa järjestyksessä kuin FASTQ-viitetiedosto.
POS	Variantin yksittäisen emäksen asema viitekromosomissa. SNP-kohteiden kohdalla tämä asema on viite-emäs yhdessä variantin kanssa; indeleiden tai poistojen suhteen tämä asema on viite-emäs välittömästi ennen varianttia.
ID (Tunnus)	Soveltuvien osien dbSNP.txt-tiedostosta saatu variantin rs-numero. Mikäli kyseisessä paikassa on useita rs-numeroita, luettelo on jaettu osiin puolipisteillä. Jos tässä asemassa ei ole dbSNP-merkintää, 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	Viiteeadista 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 laatupisteystys. 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.

## VCF-tiedostomerkinnot

Otsake	Kuvaus
<b>FILTER (SUODATIN)</b>	<p>Mikäli kaikki suodattimet hyväksytään, suodatinsarakkeeseen kirjoitetaan <b>PASS</b> (HYVÄKSYTTY).</p> <ul style="list-style-type: none"> <li>• <b>LowDP</b>— Sovelletaan paikkoihin, joiden kattavuussyvyys on alle 150x. Sellaisten amplikoniasemien kohdalla, jotka ovat sekä eteenpäin suuntautuvan että käänteisen readin kattamia, tämä vastaa 300 päällekkäistä parillisen pään readia.</li> <li>• <b>q20</b>— Laatusyvyys &lt; 20.</li> <li>• <b>MultiAllelicSite</b>— Variantti ei vastaa diploidimallia.</li> <li>• <b>R5x9</b>— Varianttintunnistusten <math>\geq 9</math> viereisten toistojen määrä (pituus 1 - 5 bp).</li> <li>• <b>SB</b>— Säikeen vinouma on annettua kynnystä suurempi.</li> </ul>
<b>INFO (TIEDOT)</b>	<p>INFO (TIEDOT) -sarakkeen mahdollisia merkintöjä ovat muun muassa seuraavat:</p> <ul style="list-style-type: none"> <li>• <b>AC</b>— Genotyypeissä oleva alleelin määrä kunkin ALT-alleelin osalta luettelussa järjestyksessä.</li> <li>• <b>AF</b>— Alleelin taajuus kunkin ALT-alleelin osalta luettelussa järjestyksessä.</li> <li>• <b>AN</b>— Alleelien kokonaismäärä tunnistetuissa genotyypeissä.</li> <li>• <b>CD</b>— Lippu, joka osoittaa, että SNP ilmenee vähintään yhden RefGene-merkinnän koodausalueella.</li> <li>• <b>DP</b>— Syvyys (aseman kanssa kohdistettujen ja varianttintunnistuksessa käytettävien emästunnistusten määrä).</li> <li>• <b>Exon</b>— Luettelo pilkulla erotettuja RefGenestä luettuja exon-alueita.</li> <li>• <b>FC</b>— Funktionaalinen seuraus.</li> <li>• <b>GI</b>— Luettelo pilkulla erotettuja RefGenestä luettuja geenitunnuksia.</li> <li>• <b>QD</b>— Variantin luottamus/laatu syvyyttä kohti.</li> <li>• <b>TI</b>— Luettelo pilkulla erotettuja RefGenestä luettuja transkriptiotunnuksia.</li> </ul>
<b>FORMAT (MUOTO)</b>	<p>Muotosarakkeessa luetellaan kaksoispisteillä erotettuja kenttiä. Esimerkiksi GT:GQ. Ilmoitettujen kenttien luettelo riippuu käytetystä varianttintunnistimesta. Saatavilla olevia kenttiä ovat muun muassa seuraavat:</p> <ul style="list-style-type: none"> <li>• <b>AD</b>— Lomakkeen X,Y, merkintä, jossa X on viitetunnistusten määrä ja Y vuoroittaisten tunnistusten määrä.</li> <li>• <b>DP</b>— Arvioitu lukusyvyys; readit, joissa MQ=255 tai huonoja pareja, suodatetaan.</li> <li>• <b>GQ</b>— Genotyypin laatu.</li> <li>• <b>GT</b>— Genotyyppi. 0 vastaa viite-emästä, 1 vastaa ALT-sarakkeen ensimmäistä merkintää jne. Eteenpäin suuntautuva vinoivi (/) osoittaa, että vaiheistustietoja ei ole saatavilla.</li> <li>• <b>NC</b>— Sellaisten emästen fraktio, joita ei tunnistettu tai joiden emäksen tunnistuksen laatu jäi minimikynnyksen alapuolelle.</li> <li>• <b>NL</b>— Melutaso; arvio emästunnistuksen melusta tässä asemassa.</li> <li>• <b>SB</b>— Säikeiden vinouma tässä asemassa. Suuremmat negatiiviset arvot viittaavat vähäisempään vinoumaan; nolaa lähellä olevat arvot viittaavat suurempaan vinoumaan.</li> <li>• <b>VF</b>— Varianttitaajuus; vuoroittaista alleelia tukevien readien prosenttiosuus.</li> </ul>
<b>SAMPLE (NÄYTE)</b>	Näytesarakkeessa ilmoitetaan FORMAT (MUOTO) -sarakkeessa määritetyt arvot.

## Genomin VCF-tiedostot

Genomin VCF (gVCF) -tiedostot ovat VCF v4.1 -tiedostoja, joissa noudatetaan tiettyjä käytänteitä kaikkien genomin sisällä olevien kohteiden edustamiseksi kohtuullisen kompaktissa muodossa.

gVCF (\*.genome.vcf.gz) -tiedostoihin sisältyvät kaikki kohdealueen kohteet yksittäisessä tiedostossa näytettä kohti.

gVCF-tiedostossa esitetään "ei tunnistuksia" asemassa, jotka eivät läpäise kaikkia suodattimia.

Genotyypin (GT) tunnisteella ./, viitataan "Ei tunnistuksia" -tilaan.

Katso lisätietoja kohdasta [sites.google.com/site/gvcftools/home/about-gvcf](https://sites.google.com/site/gvcftools/home/about-gvcf).

## Amplikonikattavuustiedosto

Amplikonikattavuustiedosto luodaan jokaisen manifestitiedoston kohdalla. Tiedostonimen M# viittaa manifestitiedoston numeroon.

Jokaisessa tiedostossa on otsikkorivi, joka sisältää manifestitiedostoon liittyvät näytetunnukset.

Tiedosto sisältää seuraavat tiedot.

- ▶ Kohdetunnus sellaisena kuin se on lueteltu manifestitiedostossa.
- ▶ Suodattimen läpäisevien readien kattavuussyvyys.

## Täydentävät tuotostiedostot

Seuraavissa tuotostiedostoissa annetaan täydentäviä tietoja tai tiivistetään ajotulokset ja analyysivirheet. Vaikka näitä tiedostoja ei tarvita analyysitulosten arviointiin, niitä voidaan käyttää vianmääritystarkoituksiin. Kaikki tiedostot sijaitsevat Alignment (Kohdistus) -kansiossa, ellei muuta ole ilmoitettu.

Tiedostonimi	Kuvaus
<b>AnalysisLog.txt</b>	Prosessointiloki, jossa kuvataan kaikki kulloisenkin ajokansion analyysin aikana suoritettavat vaiheet. Tämä tiedosto ei sisällä virheilmoituksia. Sijainti Alignment (Kohdistus) -kansiossa.
<b>AnalysisError.txt</b>	Prosessointiloki, jossa luetellaan analyysin aikana ilmenneet mahdolliset virheet. Tämä tiedosto on tyhjä, ellei virheitä ilmennyt. Sijainti Alignment (Kohdistus) -kansiossa.
<b>DemultiplexSummaryF1L1#.txt</b>	Demultiplexoinnin tulokset ilmoitetaan taulukossa, jossa on 1 rivi ruutua kohti ja 1 sarake näytettä kohti. Symboli # edustaa virtauskyvetin kaistaa 1, 2, 3 tai 4. Sijainti Alignment (Kohdistus) -kansiossa.
<b>AmpliconRunStatistics.xml</b>	Sisältää ajokohtaiset tiivistetyt tilastot. Sijainti Alignment (Kohdistus) -kansiossa.

## Analyysikansio

Analyysikansiossa on Local Run Manager -ohjelmistossa luodut tiedostot.

Tuotos- ja analyysikansion välinen suhde tiivistetään seuraavasti:

- ▶ Sekvensoinnin aikana reaaliaikainen analyysi (Real-Time Analysis I. RTA) täyttää tuotokansion kuva-analyysin, emästen tunnistamisen ja laadun pisteytyksen aikana luoduilla tiedostoilla.
- ▶ RTA kopioi tiedostot analyysikansioon reaaliajassa. Kun RTA määrää laatusihteet kunkin jakson kullekin emäkselle, ohjelmisto kirjoittaa tiedoston RTAComplete.txt molempiin kansioihin.
- ▶ Kun tiedosto RTAComplete.txt on mukana, analyysi alkaa.
- ▶ Analyysin jatkuessa Local Run Manager kirjoittaa tuotostiedostot analyysikansioon ja kopioi sitten tiedostot takaisin tuotokansioon.

## Kohdistuskansiot

Kun analyysi asetetaan uudelleen jonoon, Local Run Managerissa luodaan kohdistuskansio nimeltä **Alignment\_N**, jossa N on is a peräkkäinen numero.

## Kansion rakenne

**Alignment** (Kohdistus) —Sisältää \*.bam-, \*.vcf- ja FASTQ-tiedostoja sekä analyysimoduulikohtaisia tiedostoja.

**Date and Time Stamp** (Päivämäärä- ja kellonaikaleima)—Analyysin päivämäärä\_kellonaikaleima muodossa YYYYMMDD\_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 (Näyte1)**

- Sample1\_L001\_R1\_001\_fastq.gz

#### **Stats (Tilastot)**

- DemuxSummaryF1L1.txt
- FastqSummaryF1L1.txt

### **Data (Tiedot)**

#### **Intensities (Voimakkuudet)**

##### **BaseCalls (Emästen tunnistaminen)**

**L001** — Sisältää \*.bcl-tiedostoja.

**L001** — Sisältää \*.locs-tiedostoja.

**RTA Logs** (RTA-lokit)— Sisältää lokitiedostoja RTA-ohjelmistoanalyysistä.

**InterOp** — Sisältää binääritiedostoja, joiden avulla raportoidaan sekvensointiajon metriikkaa.

**Logs** (Lokit)— Sisältää lokitiedostoja, joissa kuvataan sekvensoinnin aikana suoritettavia vaiheita.

- RTAComplete.txt
- RunInfo.xml
- RunParameters.xml

## Emästen tunnistaminen ja indeksin monimuotoisuus

Kun näytteet sekvensoidaan NextSeq 550Dx -instrumentissa, emästen tunnistuksella määritetään emäs (A, C, G tai T) tietyn ruudun tai virtauskyvetin kuvantamisalueen jokaista klusteria kohti tietyssä jaksossa. NextSeq 550Dx -instrumentissa käytetään kaksikanavasekvensointia, joka edellyttää vain kahta kuvaa neljän DNA-emäksen koodaukseen, yhtä punaisesta ja yhtä vihreästä kanavasta.

Emäksen tunnistamisen indeksin readien prosessi eroaa emäksen tunnistamisesta muiden readien aikana.

Indeksin readien täytyy alkaa vähintään yhdellä emäksellä, joka on muu kuin G yhdessä kahdesta ensimmäisestä jaksosta. Mikäli indeksin read alkaa kahdella G-emäksen tunnistuksella, signaalin intensiteettiä ei luoda. Demultipleksoinnin suorituskyvyn varmistamiseksi signaalin on oltava läsnä jommassakummassa ensimmäisestä kahdesta jaksosta.

Kun ajon luonnin aikana valitaan indeksejä, näkyviin tulee alhaisen diversiteetin varoitus, jos indeksit eivät vastaa diversiteettivaatimuksia. Valitse alhaisen diversiteetin varoituksen välttämiseksi indeksisekvenssejä, joilla saadaan signaali molemmista kanavista kunkin jakson kohdalla.

- ▶ Punainen kanava—A tai C
- ▶ Vihreä kanava—A tai T

Tällä emästen tunnistusprosessilla varmistetaan tarkkuus kompleksiltaan matalia näytteitä analysoitaessa. Lisätietoja indeksiesi sekvensseistä on pakkausselosteessa *Mukautettu TruSeq-amplikonisarja Dx* (asiakirjanro 1000000029772).

Luotaessa ajoa Local Run Managerissa valitut testattavien näytteiden määrän. Ehdotetut indeksiyhdistelmät, jotka vastaavat indeksin diversiteettivaatimuksia, täytetään automaattisesti ohjelmistossa. Vaikka sinun ei tarvitse käyttää ehdotettuja indeksiyhdistelmiä, se on suositeltavaa.

## Versiohistoria

Asiakirja	Päivämäärä	Muutoksen kuvaus
Asiakirjanro 1000000030329 v04	Elokuu 2021	Päivitetty valtuutetun EU-edustajan osoite.
Asiakirjanro 1000000030329 v03	Joulukuu 2019	Päivitetty valtuutetun EU-edustajan osoite. Päivitetty australialaisen rahoittajan osoite.
Asiakirjanro 1000000030329 v02	Tammikuu 2019	Lisätty tietoa v2.5-reagenssisarjoista.
Asiakirjanro 1000000030329 v01	Elokuu 2018	Päivitetyt sääntelymerkinnät.
Asiakirjanro 1000000030329 v00	Marraskuu 2017	Ensimmäinen versio.



## Tekninen tuki

Teknisissä kysymyksissä voit ottaa yhteyttä Illuminan tekniseen tukeen.

Verkkosivusto: [www.illumina.com](http://www.illumina.com)  
Sähköposti: [techsupport@illumina.com](mailto:techsupport@illumina.com)

### Illuminan asiakastuen puhelinnumerot

Alue	Ilmainen	Alueellinen
Pohjois-Amerikka	+1 800 809 4566	
Alankomaat	+31 8000222493	+31 207132960
Australia	+1 800 775 688	
Belgia	+32 80077160	+32 34002973
Espanja	+34 911899417	+34 800300143
Hongkong	800960230	
Irlanti	+353 1800936608	+353 016950506
Italia	+39 800985513	+39 236003759
Itävalta	+43 800006249	+43 19286540
Japani	0800 111 5011	
Kiina	400 066 5835	
Norja	+47 800 16836	+47 21939693
Ranska	+33 805102193	+33 170770446
Ruotsi	+46 850619671	+46 200883979
Saksa	+49 8001014940	+49 8938035677
Singapore	+1 800 579 2745	
Suomi	+358 800918363	+358 974790110
Sveitsi	+41 565800000	+41 800200442
Taiwan	00806651752	
Tanska	+45 80820183	+45 89871156
Uusi-Seelanti	0800 451 650	
Yhdistynyt kuningaskunta	+44 8000126019	+44 2073057197
Muut maat	+44 1799 534000	

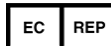
**Käyttöturvallisuustiedotteet (KTT)** – saatavilla Illuminan verkkosivustolta osoitteesta [support.illumina.com/sds.html](http://support.illumina.com/sds.html).

**Tuotedokumentaatio** – ladattavissa PDF-tiedostoina Illuminan verkkosivustolta. Siirry osoitteeseen [support.illumina.com](http://support.illumina.com), valitse tuote ja sen jälkeen **Documentation & Literature** (Tuotedokumentaatio ja -selosteet).



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**

© 2021 Illumina, Inc. Kaikki oikeudet pidätetään.

**illumina®**