

DRAGEN for Illumina DNA Prep with Enrichment Dx

NovaSeq 6000Dx tootedokumentatsioon

Muudatuste ajalugu

Dokument	Kuupäev	Muudatuse kirjeldus
200014776 v02	September 2022	Parandatud manifesti faili vorming tekstist (*.txt) BED-iks (*.bed) käituse loomise juhistes. Korrigeeritud konsensus VCF-failid VCF-failideks analüüsiväljundi jaotises.
200014776 v01	August 2022	Lisatud: Jaotis Sätted. Jaotis Süsteemne müra filtreerimine. Uuendatud käituse loomise juhiseid, et need sisaldaksid rohkem üksikasju. Parandatud kirjavead ja grammatilised vead. Täpsustatud, et juhised on mõeldud rakendusele, kui neid kasutatakse koos instrumendiga NovaSeq 6000Dx. Uuendatud VCF-i väljundfaili sisu kirjeldust.
200014776 v00	Märts 2022	Esialgne väljalase.

See dokument ja selle sisu kuuluvad ettevõttele Illumina, Inc. ja selle tütarettevõtetele („Illumina“) ning on mõeldud kasutamiseks ainult ettevõtte lepingulistele klientidele seoses selles dokumendis kirjeldatud toote (toodete) kasutamisega ega ole mõeldud mitte mingiks muuks otstarbeks. Seda dokumenti ega selle sisu ei tohi mis tahes viisil kasutada ega muul eesmärgil levitada ja/või edastada, avaldada või reprodutseerida ilma Illumina eelneva kirjaliku nõusolekuta. Illumina ei anna selle dokumendiga kolmandale isikule oma patendi-, kaubamärgi-, autori-, tava- või muu sarnase õiguse alusel mitte ühtegi litsentsi.

Kvalifitseeritud ja asjakohase koolituse saanud töötajad peavad selles dokumendis kirjeldatud juhiseid järgima rangelt ja üksikasjalikult, et tagada siin kirjeldatud toote (toodete) õige ja ohutu kasutusviis. Siinse dokumendi sisu tuleb enne nimetatud toote (toodete) kasutamist täies ulatuses läbi lugeda ja endale selgeks teha.

SELLES DOKUMENDIS KIRJELDATUD JUHISTE MITTE LUGEMINE JA MITTE ÜKSİKASJALIKULT JÄRGIMINE VÕIB KAHJUSTADA TOODET (TOOTEID), VIGASTADA INIMESI (SH KASUTAJAID VÕI TEISI) JA KAHJUSTADA MUUD VARA. NIMETATUD JUHUL EI KEHTI ÜKSKI TOOTELE (TOODETELE) ANTUD GARANTII.

ILLUMINA EI VASTUTA SELLES DOKUMENDIS KIRJELDATUD TOOTE (TOODETE) (SEALHULGAS TOOTE OSAD VÕI TARKVARA) VÄÄRKASUTUSE EEST.

© 2022 Illumina, Inc. Kõik õigused on kaitstud.

Kõik kaubamärgid kuuluvad ettevõttele Illumina, Inc. või nende vastavatele omanikele. Kaubamärgi kohta lisateabe saamiseks vt www.illumina.com/company/legal.html.

Sisukord

Muudatuste ajalugu	ii
Ülevaade	1
Analüüsimeetodid	1
Käituse loomine	4
Sätted	5
Analüüsi väljundid	7
FASTQ failid	8
BAM-failid	8
VCF-failid	9
Analüüsitulemuste kuvamine	14
Tehniline tugi	15

Ülevaade

DRAGEN™ for Illumina® DNA Prep With Enrichment Dx rakendus teostab demultipleksimise, FASTQ genereerimise, lugemi kaardistamise ja joondamise referentsgenoomiga ja variandi nimetamise, sõltuvalt valitud analüüsi töövoost.

Analüüsimeetodid

DRAGEN for Illumina DNA Prep with Enrichment Dx teostab demultipleksimise, FASTQ genereerimise, lugemi kaardistamise ja joondamise võrdlusgenoomiga olenevalt valitud töövoost:

- FASTQ genereerimine
- Geeniraja FASTQ ja VCF-i genereerimine
- Somaatiline FASTQ ja VCF-i genereerimine

FASTQ genereerimine

Koostatud järjestused kirjutatakse FASTQ failidesse proovi kohta. FASTQ failid on tekstifailid, mis sisaldavad sekveneerimisandmeid ja kvaliteediskoore ainult ühe proovi kohta. Iga proovi kohta luuakse eraldi FASTQ failid läbivooluküveti raja ja sekveneerimise lugemi kohta. Käituse seadistamisel määratletud proovi nimi lisatakse FASTQ faili nimele. FASTQ-failid on joondamise peamine sisend. FASTQ genereerimise esimene samm on demultipleksimine. Demultipleksimine määrab filtrit läbivad klastrid, proovile, võrreldes iga indeksi lugemisjärjestust käituse jaoks määratud indeksijärjestustega. Selles etapis ei arvestata kvaliteediväärtustega. Indeksite lugemid identifitseeritakse järgmiste sammudega:

- Proovid nummerdatakse alates 1-st, lähtudes nende järjestusest käituse jaoks.
- Valimi number 0 on reserveeritud klastritele, mida proovile ei määratud.
- Klastrid määratakse proovile siis, kui indeksi järjestus vastab täpselt või kui indeksi lugemi kohta on kuni üks mittevastavus.

Tarkvara sisaldab ORA tihendamist FASTQ failide tihendamiseks. ORA (*.ora) vormingu kasutamisel säilitatakse pärast tihendamise ja dekompressiooni tsüklit FASTQ sisu md5 kontrollsumma, et tagada kadudeta tihendus.

DNA kaardistamine ja joondamine

Kaardistamise esimene etapp on lugemist algete genereerimine ja siis võrdluseks kasutatavas genoomis täpsete vastete otsimine. Seejärel täiustatakse tulemusi, viies läbi täielikud Smith-Watermani joondamisprotseduurid asukohtades, kus ühtivate algete tihedus on suurim. See hästi dokumenteeritud algoritm toimib, võrreldes lugemi iga asendit kõigi referentsi kandidaatpositsioonidega. Need võrdlused vastavad lugemite ja referentsi võimalike joonduste maatriksile. Iga joonduspositsiooni variandi puhul

genereerib Smith-Watermani punktisummad, mida kasutatakse selleks, et hinnata, kas parim joendus, mis läbib maatriksraku, jõuab selleni nukleotiidivastavuse või -lahknevusega (diagonaalne liikumine), deletsiooniga (horisontaalne liikumine) või insertiooniga (vertikaalne liikumine). Lugemi ja referentsi vaste annab skoorile boonuse ning lahknevus või indel rakendab karistust. Maatriksi üldine kõrgeim punktiarvestustee on valitud joondamine.

Selle algoritmi skooride puhul valitud konkreetset väärtused näitavad, kuidas tasakaalustada, mitme võimaliku tõlgendusega joondamiseks, indeli võimalust vastandina ühele või mitmele SNP-le või kärpimata joonduse eelistusele. DRAGEN-i hindamise vaikeväärtused on mõistlikud selleks, et joondada keskmise pikkusega lugemid variandi nimetamise rakendused terve inimese referentsoomiga. Smith-Watermani hindamisparameetrite komplektid on genoomilise mutatsiooni ja sekveneerimisvigade ebatäpne mudel. Mõne rakenduse puhul võivad sobida erinevalt häälestatud joondamishindamisväärtused.

DRAGEN geeniraja variandi nimetamine

DRAGEN geeniraja väikesemahuline variandi nimetamine võtab sisendiks kaardistatud ja joondatud DNA lugemid ning nimetab SNP-d ja indelid, kasutades haplotüüpide kohaliku *de novo* tuvastuse ja veergudevahelise tuvastuse kombinatsiooni.

Nimetatavad võrdluspiirkonnad tuvastatakse kõigepealt piisava joonduse katvusega. Nendes võrdluspiirkondades identifitseerib sortitud lugemite kiire skannimine aktiivsed piirkonnad, mis on tsentreeritud korrelatsiooniveergude ümber, kus esineb tõendeid variandi kohta. Aktiivsed piirkonnad on polsterdatud piisava kontekstiga, et katta lähedalasuv oluline mittevõrdlussisu. Kui on tõendeid indelide kohta, saavad aktiivsed piirkonnad täiendavat polstrit.

Joondatud lugemid kärbitakse igasse aktiivsesse piirkonda ja koondatakse De Bruijni graafikule. Kärbitud lugemite servad kaalutakse tähelepanekute arvu põhjal, mille aluseks on võrdlusjärjestus. Pärast mõningast graafiku puhastamist ja lihtsustamist eraldatakse kõik allika ja sisendi vahelised teed kandidaathaplotüüpideks. Iga haplotüüp on Smith-Waterman, mis on joondatud võrdlusgenoomiga, et tuvastada esindatavad variandid. Seda sündmuste kogumit võib täiendada asukohapõhise tuvastamisega. Iga lugemihaplotüübi paari puhul hinnatakse lugemi jälgimise tõenäosust $P(r|H)$, eeldades, et haplotüüp on tõeline algproov, kasutades paaris varjatud Markovi mudelit (HMM).

Kandidaatide genotüübid moodustatakse referentsasendi järgi aktiivses piirkonnas variantsündmuste (SNP-d või indelid) diploidsetest kombinatsioonidest. Iga juhtumi (sh referentsi) puhul hinnatakse iga kattuva lugemi tingimusliku tähelepaneku tõenäosust $P(r|e)$ sündmust toetavate haplotüüpide hinnangulise maksimaalse $P(r|H)$ väärtusena. Need liidetakse tingimuslikuks tõenäosuseks $P(r|e1e2)$ genotüübi puhul (sündmuste paar) ja korrutatakse, et saada tingimuslik tõenäosus $P(R|e1e2)$ kogu lugemikogumite tähelepaneku tekkeks. Bayesi valemi abil arvutatakse iga diploidse genotüübi tagumine tõenäosus $P(e1e2|R)$ ja võitja nimetatakse.

gVCF-režiimis, mida kasutatakse skaleeritava mitme proovi variandi nimetamiseks, saab DRAGEN Germline Small Variant Caller rakendust kasutada proovi kohta, et luua vahepealne genoomse variandi

nimetuse fail (gVCF). gVCF-i saab seejärel kasutada mitme proovi efektiivseks ühiseks genotüübi määramiseks, mis võimaldab proovide kiiret inkrementaalset töötlemist ja skaleerimist suurte kohortideni.

Kuna DRAGEN Germline Small Variant Caller rakendusel on algoritmid, mis võimaldavad tõhusalt eristada korrelatsioonivigu tegelikest variantidest, on filtreerimisreeglid väga lihtsad.

DRAGEN Somatic Variant Calling

DRAGEN Somatic Small Variant Caller võtab kaardistatud ja joondatud DNA lugemid sisendina ning nimetab SNV-d ja indelid haplotüüpide kohaliku *de novo* kooste kaudu aktiivses piirkonnas.

Nimetatavad võrdluspiirkonnad tuvastatakse kõigepealt piisava joonduse katvusega. Nendes võrdluspiirkondades identifitseerib sorteeritud lugemite skaneering aktiivseid piirkondi, mis on tsentreeritud kuhja veergude ümber, kus esineb tõendeid kasvaja lugemite variandi kohta. Aktiivsed piirkonnad on polsterdatud piisava kontekstiga, et katta lähedalasuv oluline mittevõrdlussisu. Kui on tõendeid indelide kohta, saavad aktiivsed piirkonnad täiendavat polstrit.

Joondatud lugemid kärbitakse igasse aktiivsesse piirkonda ja koondatakse De Bruijni graafikule. Kärbitud lugemite servad kaalutakse tähelepanekute arvu põhjal, mille aluseks on võrdlusjärjestus. Pärast mõningast graafiku puhastamist ja lihtsustamist eraldatakse kõik allika ja sisendi vahelised teed kandidaathaplotüüpideks. Iga haplotüüp on Smith-Waterman, mis on joondatud võrdlusgenoomiga, et tuvastada esindatavad variandid. Iga lugemi haplotüübi paari puhul hinnatakse lugemi jälgimise tõenäosust $P(r|H)$, kasutades paaris varjatud Markovi mudelit (HMM), eeldades, et haplotüüp on tõeline algproov.

TLOD-skoori määramiseks skannib DRAGEN Somatic Small Variant Caller esmalt võrdluspositsiooni iga somaatilise sündmuse ja võrdlussündmuse järgi aktiivses piirkonnas. Tingimuslik tõenäosus $P(r|e)$ iga kattuva lugemi jälgimise kohta hinnatakse sündmust toetavate haplotüüpide maksimaalse $P(r|H)$ arvuna. Need liidetakse sündmuse hüpoteesi (E) tingimuslikuks tõenäosuseks $P(r|E)$, mis hõlmab võrdluse ja kandidaat-somaatilise alleeli segu mitmetel erinevatel võimalikel alleelisagedustel, mis seejärel korrutatakse, et saada tingimuslik tõenäosus $P(R|E)$ kogu lugemite kogumi jälgimiseks. Järgnevalt arvutatakse TLOD-skoor tõendina selle kohta, et kasvajaproovis esineb antud lookuses ALT-alleel.

Käituse loomine

Kasutage järgmisi samme käituse seadistamiseks rakenduses Illumina käivitushaldur seadmes NovaSeq 6000Dx või kasutades võrguühendusega arvuti brauserit. Proovi andmeid saab sisestada käsitsi või proovilehe importimise teel.

Rakenduse ja käituse sätted

1. Valige käituste kuvalt **Create Run** (Loo käitus).
2. Valige rakendus DRAGEN for Illumina DNA Prep with Enrichment Dx ja valige seejärel **Next** (Järgmine).
3. Sisestage kuval Run Settings (Käituse sätted) käituse nimi. Käituse nimi identifitseerib käituse sekveneerimisest analüüsini.
4. **[Valikuline]** Sisestage käituse kirjeldus käituse täiendavaks tuvastamiseks.
5. Veenduge, et valitud teegi ettevalmistuskomplekt oleks Komplekt Illumina DNA Prep With Enrichment Dx teegi ettevalmistuskomplekt.
6. Valige soovitud indeksadapteri komplekt.
7. Sisestage lugemi pikkus.
Lugem 1 ja lugem 2 on vaikeväärtusega 151 tsükli.
Indeks 1 ja indeks 2 on fikseeritud väärtusega 10 tsükli.
8. **[Valikuline]** Sisestage teegikatsuti ID.
9. Valige **Next** (Edasi).

Proovi andmed

Proovi teabe käsitsi sisestamiseks kasutage kuva Sample Data (Proovi andmed) tabelit. Proovi teabe üleslaadimiseks valige alternatiivselt **Import Samples** (Proovide importimine). Teavet proovide teabe importimise kohta vaadake jaotisest [Proovide importimine leheküljel 5](#) (Proovide importimine).

Proovide sisestamine käsitsi

1. Sisestage väljale Sample ID (Proovi ID) unikaalne proovi ID.
2. Süvendi asendi valimiseks kasutage suvandit **Plate – Well Position** (Plaat – süvendi asend). Väljad i7 Index, Index 1, i5 Index ja Index 2 täidetakse automaatselt.
3. **[Valikuline]** Sisestage teegi nimi.
4. Lisage ridu ja korrake samme **1–3** vastavalt vajadusele, kuni kõik proovid on tabelisse lisatud.
5. Valige **Next** (Edasi).

Proovide importimine

Kui planeerite seadme Illumina käivitushaldur kätust võrguühendusega arvutis veebilehitsejat kasutades, on prooviandmete kuvall allaadimiseks saadaval mall (*.csv).

1. Tühja CSV-faili allalaadimiseks valige käsk **Download Template** (Laadi mall alla).
2. Sisestage CSV-faili proovi andmed ja salvestage fail.
Proovilehe CVS-fail sisaldab järgmisi andmeveerge: Proovi ID, plaat – süvendi asend, **valikuline** teegi nimi.
3. Valige suvand **Import Samples** (Impordi proovid) ja sirvige CSV-faili asukohta.
4. Valige **Next** (Edasi).

Analüüsi sätted

1. Valige soovitud analüüsi töövoog.
 - FASTQ genereerimine
 - Germline FASTQ ja VCF-i genereerimine geeniraja töövoog jaoks
 - Somaatiline FASTQ ja VCF-i genereerimine somaatilise töövoog jaoks
2. **[Valikuline]** Soovi korral märkige ruut **Generate ORA compressed FASTQs** (Genereeri ORA tihendusega FASTQ-d), et lubada FASTQ ORA tihendamine.
3. **[VCF-i genereerimise töövoog]** Kasutage manifestifaili valimiseks rippmenüüd **Manifest File Selection** (Manifestifaili valimine).
Rakenduse DRAGEN for Illumina DNA Prep with Enrichment Dx jaoks on manifestifail kohustuslik sisend. Manifest on tabeldusmärgiga eraldatud BED (*.bed) fail, mis määratleb sihtreferentspiirkondade nimed ja asukohad.
4. **[Somaatiline FASTQ ja VCF genereerimise töövoog]** Mürafaili valimiseks kasutage rippmenüüd **Noise File Selection** (Mürafaili valimine).
Süsteemse müra filtreerimiseks saab määrata asukohaspetsiifilise müratasemega BED-faili. Lisateavet leiate jaotisest [Müra filtreerimine leheküljel 6](#).
5. Valige **Next** (Edasi).

Analüüs Ülevaade

1. Vaadake üle kuvall Review (Ülevaatus) kuvadel Run Settings (Käituse sätted), Sample Data (Prooviandmed) ja Analysis Settings (Analüüsi sätted) sisestatud teave.
2. Valige käsk **Save** (Salvesta).
Käitus salvestatakse kuva Runs (Käitused) vahekaardile Planned (Planeeritud).

Sätted

Valige rakendus rakenduste kuvall, et vaadata praegusi sätteid ja muuta sätteid.

Konfiguratsioon

Konfiguratsioonikuval kuvatakse järgmised rakenduse sätted.

- **Library Prep Kits (Teegi ettevalmistuskomplektid)**– kuvab rakenduse jaoks vaikimisi teegi ettevalmistuskomplekti. Seda sätet ei saa muuta.
- **Index Adapter Kits (Indeksadapteri komplektid)**– kuvab rakenduse vaikimisi kasutatava indeksi adapteri komplekti. Seda sätet ei saa muuta.
- **Read lengths (Lugemipikkused)**– vaikimisi seatakse rakenduse lugemipikkusteks 151, kuid neid saab muuta käituse loomisel.
- **Manifest and Noise Files (Manifesti- ja mürafailid)**– manifesti- ja mürafailide sätete üleslaadimine ja muutmine.
 - Valige **Upload File** (Laadi fail üles), et laadida üles failid analüüsis kasutamiseks.
 - Valige raadionupp **Default** (Vaikimisi), et määrata fail käituse loomisel valitud vaikimisi manifesti- või mürafailiks, kui rakendus on valitud.
 - Tähistage märkeruut **Enabled** (Lubatud), et seada fail, mida kuvatakse rippmenüüs käituse loomisel.

Õigused

Kasutage märkeruute õiguste kuval, et hallata rakenduse kasutajate juurdepääsu.

Müra filtreerimine

Süsteemne müra filtreerimine on saadaval, kui kasutate somaatilist töövoogu. Filtrit saab kasutada režiimis Tumor-Normal (Tuumor - normaalne), kuid see on eriti kasulik Tumor-Only (Tuumor – ainult) käitustes, mille puhul pole olemas normaalselt sobitatud tulemusi.

Tavalistest proovidest tuleb genereerida süsteemse müra BED. Soovitav on luua süsteemsed mürafailid, mis on spetsiifilised teegi ettevalmistusele, sekveneerimissüsteemile ja paneelile. Mürafaili loomiseks on soovitatav kasutada ligikaudu 50 tavalist proovi.

Analüüsi väljundid

DRAGEN for Illumina DNA Prep with Enrichment Dx salvestab analüüsikausta järgmised andmed. PDF-i annavad ainult geeniraja ja somaatilised töövood.

- Kasutatud manifestifail
- Tarkvaraversioon
- Proovi ID-d
- Joondatud lugemite koguarv
- Ühtlustatud lugemite osakaal proovi kohta
- Proovi kohta nimetatud SNV-de arv
- Proovi kohta nimetatud indelide arv
- Katvuse statistika

Analüüsi väljundfailid

Rakendus loob järgmised väljundfailid. Täpsed loodavad failid sõltuvad kasutatavast analüüsi töövoost. Väljundfailid asuvad analüüsi kaustas.

Väljundfail	Kirjeldus
FASTQ (*.fastq.gz või *.fastq.ora)	Vahefailid, mis sisaldavad kvaliteediskooriga aluste määramisi. FASTQ-failid on joondamisetapi peamine sisend. Kui valitud on ORA tihendamine, siis kajastub see faili nimes.
Joondamise BAM-failid (*.bam)	Sisaldab antud proovi joondatud lugemeid.
Genoomi VCF-failid (*.gvcf.gz)	Sisaldab iga positsiooni jaoks genotüüpi, olenemata sellest, kas seda nimetatakse variandiks või etaloniks.
VCF-failid (*.vcf.gz)	Sisaldab igas positsioonis nimetatud variante.
Käituse mõõdikute raport (*.csv)	Sisaldab kvaliteedimõõdikuid käituse kohta, sh kogutootlus ja Q30 skoor.

FASTQ failid

FASTQ (*.fastq.gz, *.fastq.ora) on tekstipõhine failivorming, mis sisaldab aluste nimetamisi ja kvaliteedi väärtusi ühe lugemi kohta. Iga fail sisaldab järgmist teavet.

- Proovi identifikaator
- Järjestus
- Plussmärk (+)
- Phredi kvaliteediskoorid ASCII + 33 kodeeritud vormingus

Proovi identifikaator vormindatakse järgmiselt.

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y
ReadNum:FilterFlag:0:SampleNumber
Näide:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAA9#:<#<;<<<????#=#
```

BAM-failid

BAM-fail (*.bam) on SAM-faili (sekveneerimise joonduse kaardi) tihendatud binaarversioon, mida kasutatakse kuni 128 Mb joondatud järjestuste kuvamiseks. BAM-failid kasutavad failinimede vormingut `SampleName_S#.bam`, kus # on proovi number, mis on määratud käituseks järjestatud proovide järjekorral. Mitme sõlme režiimis on S# seatud S1-le, olenemata proovi järjekorrast.

BAM-failid sisaldavad päisejaotist ja joondusejaotist:

- Päis – sisaldab teavet kogu faili kohta, näiteks proovi nimi, proovi pikkus ja joondusmeetod. Joondusejaotises seostatakse joondused päisejaotises oleva konkreetse teabega.
- Joondused – sisaldab lugemi nime, lugemi järjestust, lugemi kvaliteeti, joondamisteavet ja kohandatud silte. Lugemi nimi sisaldab kromosoomi, alguskoordinaati, joondamiskvaliteeti ja vastavuse deskriptorstringi.

Joondusejaotises on iga lugemi või lugemipaari kohta järgmine teave:

- AS: Paaristulemusega joondamiskvaliteet.
- RG: Lugemirühm, mis näitab konkreetse proovi lugemite arvu.
- BC: Vöötkoodi silt, mis näitab lugemiga seotud demultipleksitud proovi ID-d.
- SM: Üksiktulemusega joondamiskvaliteet.
- XC: Vastavuse deskriptorstring

- XN: Amplikoni nimetuse silt, mis salvestab lugemi BAM indeksfailidega (*.bam.bai) seotud amplikoni ID, mis annab vastava BAM-faili indeksi.

VCF-failid

Variandi nimetamise formaadi (*.vcf) failid sisaldavad teavet variantide kohta, mis on leitud referentsgenoomi kindlatel positsioonidel.

VCF-faili päis sisaldab VCF-failivormingu versiooni ja variandi nimetaja versiooni ning loetleb faili ülejäänud osas kasutatud annotatsioonid. VCF-i päis sisaldab ka referentsgenoomi faili ja BAM-faili. Päise viimane rida sisaldab andmeridade veerupäiseid. Iga VCF-faili andmerida sisaldab teavet ühe variandi kohta.

tabel 1 VCF-faili pealkirjad

Pealkiri	Kirjeldus
CHROM	Referentsgenoomi kromosoom. Kromosoomid ilmuvad samas järjekorras kui viitefailis FASTA.
POS	Variandi ühe-aluseline positsioon referentskromosoomis. Ühe nukleotiidi variantide (SNV-d) puhul on see positsioon variandiga referentsaluseks. Indelite puhul on see positsioon referentsalus vahetult enne varianti.
ID	Rakenduvuse korral SNP rs number (referentsi SNP), mis on hangitud failist dbSNP.txt. Kui selles asukohas on mitu rs-numbrit, on loend semikoolonitega eraldatud. Kui sellel positsioonil ei ole dbSNP-kirjet, kasutatakse puuduva väärtuse tähist ('.').
REF	Referentsgenotüüp. Näiteks ühe T kustutamine on kujutatud referentsina TT ja alternatiivina T. A-T üksiknukleotiidi variant on kujutatud referentsina A ja alternatiivina T.
ALT	Alleelid, mis erinevad referentslugemist. Näiteks on üksiku T sisestamine esindatud kui referents A ja alternatiiv AT. Üksiknukleotiidi variant A-T on esindatud kui referents A ja alternatiiv T.
QUAL	Phred-skaalaga kvaliteediskoor, mille määrab variandi nimetaja. Kõrge skoor näitab suuremat usaldust variandi suhtes ja väiksemat vigade tõenäosust. Q kvaliteediskoori korral on vea hinnanguline tõenäosus 10 ⁻ (Q/10). Näiteks on Q30 nimetusekomplekti veamäär 0,1%. Paljud variantide nimetajad määravad oma statistiliste mudelite põhjal kvaliteediskoori, mis on täheldatud veamäära suhtes kõrge.

tabel 2 VCF-faili annotatsioonid

Pealkiri	Kirjeldus
FILTER	<p>Kui kõik filtrid on läbitud, kirjutatakse filtriveergu PASS.</p> <p>Geeniraja töövoos võimalikud FILTER sissekanded on järgmised:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL– rakendatakse, kui SNP variandi QUAL skoor ei vasta lävele • DRAGENIndelHardQUAL– rakendatakse juhul, kui indeli variandi QUAL skoor ei vasta lävele • LowDepth– sait on filtreeritud, sest kattesügavus ei vasta lävele • LowGQ– sait on filtreeritud, sest genotüübi kvaliteet ei vasta lävele • PloidyConflict– genotüübi nimetamine variandi nimetajalt, ei ole kooskõlas kromosoomi ploidsusega • base_quality– sait on filtreeritud, sest selle lookuse alt-lugemite mediaanne baaskvaliteet ei vasta lävele • filtered_reads– sait on filtreeritud, sest liiga suur osa lugemeid on välja filtreeritud • fragment_length– sait filtreeritud, sest kõigi alt-lugemite fragmendi mediaanpikkuste ja referentslugemite mediaanpikkuste absoluutne erinevus antud lookuses ületab läve • low_depth– sait filtreeritud, sest lugemi sügavus on liiga madal • low_frac_info_reads– sait on filtreeritud, sest informatiivsete lugemite fraktsioon on allpool läve • low_normal_depth– sait on filtreeritud, sest normaalne proovi lugemi sügavus on liiga madal • long_indel– sait on filtreeritud, kuna indel on liiga pikk • mapping_quality– sait filtreeritud, sest selle lookuse alt-lugemite mediaanne kaardistus kvaliteet ei vasta lävele • multiallelic– sait filtreeritud, sest rohkem kui kaks alt alleeli läbivad kasvaja LOD-i • non_homref_normal– sait filtreeritud, sest normaalne proovi genotüüp ei ole homosügootne referents • no_reliable_supporting_read– sait filtreeritud, sest usaldusväärset soomaatilist lugemid ei ole • panel_of_normals– kuvatakse vähemalt ühes proovis normaalväärtuste paneelis vcf • read_position– sait on filtreeritud, sest lugemi alguse/ lõpu ja selle lookuse vahemaade mediaan on alla läve • RMxNRepeatRegion– sait on filtreeritud, sest kogu või osa variantalleelist on referentsväärtuse kordus • strand_artifact– sait filtreeritud tugeva ahela hälbe tõttu • str_contraction– sait filtreeritud kahtlustatava PCR-i vea tõttu, kus alt-alleel on ühe kordusüksuse võrra referentsist väiksem • too_few_supporting_reads– sait on filtreeritud, sest kasvaja proovis on liiga vähe toetavaid lugemeid • weak_evidence– soomaatilise variandi skoor ei vasta lävele

Pealkiri	Kirjeldus
FILTER (jätkub)	<p>Somaatiliste töövoogude võimalikud FILTRI kirjed on järgnevad:</p> <ul style="list-style-type: none"> • base_quality– sait on filtreeritud, sest selle lookuse alt-lugemite mediaanne baaskvaliteet ei vasta lävele • filtered_reads– sait filtreeritud, sest liiga suur osa lugemeid on väljafiltreeritud • fragment_length– sait filtreeritud, sest kõigi alt-lugemite fragmendi mediaanpikkuste ja referentslugemite mediaanpikkuste absoluutne erinevus antud lookuses ületab läve • low_depth– sait filtreeritud, sest lugemi sügavus on liiga madal • low_frac_info_reads– sait on filtreeritud, sest informatiivsete lugemite fraktsioon on allpool läve • low_normal_depth– sait on filtreeritud, sest normaalne proovi lugemi sügavus on liiga madal • long_indel– sait on filtreeritud, kuna indel on liiga pikk • mapping_quality– sait filtreeritud, sest selle lookuse alt-lugemite mediaanne kaardistus kvaliteet ei vasta lävele • multiallelic– sait filtreeritud, sest rohkem kui kaks alt alleeli läbivad kasvaja LOD-i • non_homref_normal– sait filtreeritud, sest normaalne proovi genotüüp ei ole homosügootne referents • no_reliable_supporting_read– sait filtreeritud, sest usaldusväärset somaatilist lugemid ei ole • panel_of_normals– kuvatakse vähemalt ühes proovis normaalväärtuste paneelis vcf • read_poision– sait on filtreeritud, sest lugemi alguse/ lõpu ja selle lookuse vahemaade mediaan on alla läve • RMxNRepeatRegion– sait on filtreeritud, sest kogu või osa variantalleelist on referentsväärtuse kordus • strand_artifact– sait filtreeritud tugeva ahela hälbe tõttu • str_contraction– sait filtreeritud kahtlustatava PCR-i vea tõttu, kus alt-alleel on ühe kordusüksuse võrra referentsist väiksem • too_few_supporting_reads– sait on filtreeritud, sest kasvajaproovis on liiga vähe toetavaid lugemeid • weak_evidence–somaatilise variandi skoor ei vasta lävele • systematic_noise– sait filtreeritud lähtuvalt süsteemse müra tõenditest normaalväärtustel

Pealkiri	Kirjeldus
TEAVE	<p>Geeniraja töövoo võimalikud INFO kirjed on järgmised.</p> <ul style="list-style-type: none"> • AC– alleelide arv genotüüpides iga ALT-alleeli jaoks, samas järjekorras nagu loetletud. • AF– alleelide sagedus iga ALT-alleeli jaoks, samas järjekorras nagu loetletud. • AN– alleelide koguarv nimetatud genotüüpides. • DB– dbSNP liikmesus. • FS– Phred-skaalaga p-väärtus, kasutades Fisheri täpset testi ahela hälbe tuvastamiseks. • QD– variandi usaldatavus/kvaliteet sügavuse järgi. • R2_5P_bias– skoor, mis põhineb mate kõrvalekaldel ja vahekaugusel ahela 5' otsas. • SOR– sümmeetriline tõenäosuste suhe 2x2 kontingentsustabelis ahela hälbe tuvastamiseks. • DP– ligikaudne lugemisügavus (informatiivne ja mitteinformatiivne); mõned lugemid võivad olla filtreeritud mapq jms alusel. • END– intervalli peatamisasend. • FractionInformativeReads– informatiivsete lugemite osa lugemite koguarvust. • MQ– RMS-i kaardistamise kvaliteet. • MQRankSum– Wilcoxon astaksummatesti Z-skoor Alt vs Ref lugemi kaardistamise kvaliteedinäitajatest. • ReadPosRankSum– Wilcoxon astaksummatesti Z-skoor Alt vs Ref lugemi positsiooni hälbest. • SOMATIC– vähemalt üks variant selles positsioonis on somaatiline. <p>Somaatilise töövoo võimalikud INFO kirjed on järgnevad:</p> <ul style="list-style-type: none"> • DP– ligikaudne lugemisügavus (informatiivne ja mitteinformatiivne); mõned lugemid võivad olla filtreeritud mapq jms alusel. • END– intervalli peatamisasend. • FractionInformativeReads– informatiivsete lugemite osa lugemite koguarvust. • MQ– RMS-i kaardistamise kvaliteet. • MQRankSum– Wilcoxon astaksummatesti Z-skoor Alt vs Ref lugemi kaardistamise kvaliteedinäitajatest. • ReadPosRankSum– Wilcoxon astaksummatesti Z-skoor Alt vs Ref lugemi positsiooni hälbest. • AQ– süsteemimüra skoor. • hotspot– tuntud somaatiline sait, mida kasutatakse kindluse suurendamiseks nimetamisel. • SOMATIC– vähemalt üks variant selles positsioonis on somaatiline.

Pealkiri	Kirjeldus
VORMING	<p>Vormindatud veerus on koolonitega eraldatud väljad. Nt GT:GQ. Geeniraja töövoos saadaolevad väljad on järgmised.</p> <ul style="list-style-type: none"> • AD– alleelsügavused (ainult informatiivsed lugemid kõigist lugemitest) loetletud järjestuses ref ja alt alleelide kohta. • AF– alleelifraktsioonid alt alleelide kohta loetletud järjekorras. • DP– ligikaudne lugemisügavus (lugemid MQ = 255 või halbade paarilistega filtreeritakse). • F1R2– F1R2 paari orientatsiooniga lugemite arv, mis toetab iga alleeli. • F2R1– F2R1 paari orientatsiooniga lugemite arv, mis toetab iga alleeli. • GP– Phred-skaalaga posterioorsed tõenäosused VCF-i spetsifikatsioonis määratletud genotüüpide puhul. • GQ– genotüübi kvaliteet. • GT– genotüüp. 0 vastab etalonlusele, 1 vastab veeru ALT esimesele kirjele ja nii edasi. Kalkkriips (/) näitab, et faasimise teave pole saadaval. • MB– proovikomponendi statistika mate hälbe tuvastamiseks. • PL– normaliseeritud, Phred-skaalaga tõenäosus genotüüpidele vastavalt VCF-i spetsifikatsiooni määratlusele. • PRI– Phred-skaalaga eelnevad tõenäosused genotüüpide suhtes. • PS– füüsilise faasimise ID-teave, kus iga unikaalne ID konkreetsetes proovis (kuid mitte proovide vahel) ühendab kirjeid faasimise rühmas. • SB– proovikomponendi statistika, mis sisaldab Fisher'i täpset testi ahela hälbe tuvastamiseks. • SQ– somaatiline kvaliteet. <p>Somaatiline töövoos saadaval olevad väljad on järgmised.</p> <ul style="list-style-type: none"> • AD– alleelsügavused (ainult informatiivsed lugemid kõigist lugemitest) loetletud järjestuses ref ja alt alleelide kohta. • AF– alleelifraktsioonid alt alleelide kohta loetletud järjekorras. • DP– ligikaudne lugemisügavus (lugemid MQ = 255 või halbade paarilistega filtreeritakse). • F1R2– F1R2 paari orientatsiooniga lugemite arv, mis toetab iga alleeli. • F2R1– F2R1 paari orientatsiooniga lugemite arv, mis toetab iga alleeli. • GT– genotüüp. 0 vastab etalonlusele, 1 vastab veeru ALT esimesele kirjele ja nii edasi. Kalkkriips (/) näitab, et faasimise teave pole saadaval. • MB– proovikomponendi statistika mate hälbe tuvastamiseks. • PS– füüsilise faasimise ID-teave, kus iga unikaalne ID konkreetsetes proovis (kuid mitte proovide vahel) ühendab kirjeid faasimise rühmas. • SB– proovikomponendi statistika, mis sisaldab Fisher'i täpset testi ahela hälbe tuvastamiseks. • SQ– somaatiline kvaliteet.
PROOV	Proovi veerg annab veerus VORMING määratud väärtused.

Genoomi VCF-failid

Genoomi VCF-failid (*.gvcf.gz) järgivad genoomi kõigi saitide mõistlikult kompaktses vormis esitamise tavasid. GVCF failid sisaldavad ühes failis kõiki huvipakkuvat piirkonnas asuvaid saite iga proovi kohta. GVCF-fail näitab nimetuse puudumist positsioonidel, mis ei läbi kõiki filtreid. Genotüübi (GT) silt ./ näitab nimetuse puudumist.

Analüüsitulemuste kuvamine

Käimasolevad käitused kuvatakse vahekaardil Active (Aktiivne). Lõpetatud käitused kuvatakse vahekaardil Completed (Lõpetatud). Lisateavet tulemuste vaatamise kohta vt jaotisest [NovaSeq 6000Dx toote dokumentatsioon \(dokumendi nr 200010105\)](#).

Tehniline tugi

Tehnilise abi saamiseks pöörduge Illumina tehnilise toe poole.

Veebisait: www.illumina.com
E-post: techsupport@illumina.com

Illumina tehnilise toe telefoninumbrid

Piirkond	Tasuta	Rahvusvaheline
Austraalia	+61 1800 775 688	
Austria	+43 800 006249	+43 1 9286540
Belgia	+32 800 77 160	+32 3 400 29 73
Kanada	+1 800 809 4566	
Hiina		+86 400 066 5835
Taani	+45 80 82 01 83	+45 89 87 11 56
Soome	+358 800 918 363	+358 9 7479 0110
Prantsusmaa	+33 8 05 10 21 93	+33 1 70 77 04 46
Saksamaa	+49 800 101 4940	+49 89 3803 5677
Hongkong, Hiina	+852 800 960 230	
India	+91 800 6500 375	
Indoneesia		00 7803 651 0048
Iirimaa	+353 1800 936608	+353 1 695 0506
Itaalia	+39 800 985513	+39 236 003 759
Jaapan	+81 0800 111 5011	
Malaisia	+60 1800 80 6789	
Holland	+31 800 022 2493	+31 20 713 2960
Uus-Meremaa	+64 800 451 650	
Norra	+47 800 16 836	+47 21 93 96 93
Filipiinid	+63 1800 1651 0798	
Singapur	1 800 5792 745	
Lõuna-Korea	+82 80 234 5300	
Hispaania	+34 800 300 143	+34 911 899 417

Piirkond	Tasuta	Rahvusvaheline
Rootsi	+46 2 0088 3979	+46 8 5061 9671
Šveits	+41 800 200 442	+41 56 580 00 00
Taiwan, Hiina	+886 8 0665 1752	
Tai	+66 1800 011 304	
Ühendkuningriik	+44 800 012 6019	+44 20 7305 7197
Ameerika Ühendriigid	+1 800 809 4566	+1 858 202 4566
Vietnam	+84 1206 5263	

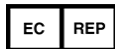
Ohutuse andmelehed (safety data sheets, SDS-id): saadaval Illumina veebisaidil support.illumina.com/sds.html.

Toote dokumendid: allalaaditavad veebisaidilt support.illumina.com.



Illumina
5200 Illumina Way
San Diego, California 92122 USA
+1 800 809 ILMN (4566)
+1 85 8202 4566 (väljaspool Põhja-Ameerikat)
techsupport@illumina.com
www.illumina.com

CE



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

Sponsor Austraalias

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

KASUTAMISEKS IN VITRO DIAGNOSTIKAS

© 2022 Illumina, Inc. Kõik õigused on kaitstud.

illumina[®]