

DRAGEN for Illumina DNA Prep with Enrichment Dx

Productdocumentatie voor NovaSeq 6000Dx

Revisiegeschiedenis

| Document | Datum | Omschrijving van wijziging |
|---------------|----------------|--|
| 200014776 v02 | September 2022 | Indeling voor manifestbestand gecorrigeerd van tekst (*.txt) naar BED (*.bed) in instructies voor het maken van een run. Consensus VCF-bestanden gecorrigeerd naar VCF-bestanden in de paragraaf over analyse-uitvoer. |
| 200014776 v01 | Augustus 2022 | Toegevoegd: Sectie Instellingen. Sectie Systematische ruisfiltering. Instructies voor het maken van runs bijgewerkt om meer details op te nemen. Typ- en grammaticafouten verbeterd. Vermeld dat de instructies bedoeld zijn voor de app bij gebruik met het NovaSeq 6000Dx-instrument. Informatie bijgewerkt over de inhoud van VCF-uitvoerbestanden. |
| 200014776 v00 | Maart 2022 | Eerste uitgave. |

Dit document en de inhoud ervan zijn eigendom van Illumina, Inc. en haar dochterondernemingen ('Illumina') en zijn alleen bedoeld voor contractueel gebruik door haar klanten in verband met het gebruik van het/de hierin beschreven product(en) en voor geen enkel ander doel. Dit document en de inhoud ervan mogen niet zonder de voorafgaande schriftelijke toestemming van Illumina worden gebruikt of gedistribueerd voor welk ander doel dan ook en/of op een andere manier worden gecommuniceerd, geopenbaard of gereproduceerd. Illumina verleent met dit document geen licenties onder zijn octrooi-, handelsmerk-, auteursrecht of gewoonterechten noch soortgelijke rechten van derden.

De instructies in dit document moeten strikt en uitdrukkelijk worden opgevolgd door gekwalificeerd en voldoende opgeleid personeel om een correct en veilig gebruik van het/de hierin beschreven product(en) te waarborgen. Alle inhoud van dit document moet volledig worden gelezen en begrepen voordat dergelijk(e) product(en) worden gebruikt.

HET NIET VOLLEDIG LEZEN EN UITDRUKKELIJK OPVOLGEN VAN ALLE INSTRUCTIES IN DIT DOCUMENT KAN RESULTEREN IN SCHADE AAN DE PRODUCTEN, LETSEL AAN PERSONEN (INCLUSIEF GEBRUIKERS OF ANDEREN) EN SCHADE AAN ANDERE EIGENDOMMEN. BIJ HET NIET VOLLEDIG LEZEN EN UITDRUKKELIJK OPVOLGEN VAN ALLE INSTRUCTIES IN DIT DOCUMENT VERVALLEN ALLE GARANTIES DIE VAN TOEPASSING ZIJN OP HET/DE PRODUCT (EN).

ILLUMINA IS OP GEEN ENKELE MANIER AANSPRAKELIJK VOOR GEVOLGEN VAN EEN ONJUIST GEBRUIK VAN DE PRODUCTEN DIE HIERIN WORDEN BESCHREVEN (INCLUSIEF DELEN DAARVAN OF SOFTWARE).

© 2022 Illumina, Inc. Alle rechten voorbehouden.

Alle handelsmerken zijn het eigendom van Illumina, Inc. of van hun respectievelijke eigenaren. Ga naar www.illumina.com/company/legal.html voor meer informatie over specifieke handelsmerken.

Inhoudsopgave

| | |
|---------------------------------------|-----------|
| Revisiegeschiedenis | ii |
| Overzicht | 1 |
| Analysemethoden | 1 |
| Een run aanmaken | 5 |
| Instellingen | 7 |
| Analyse-uitvoerbestanden | 8 |
| FASTQ-bestanden | 9 |
| BAM-bestanden | 9 |
| VCF-bestanden | 10 |
| Analyseresultaten weergeven | 16 |
| Technische ondersteuning | 17 |

Overzicht

De DRAGEN™ voor Illumina® DNA Prep with Enrichment Dx-toepassing voert demultiplexing, FASTQ-bestanden genereren, bepalingstoewijzing en uitlijning naar een referentiegenoom en variantbepaling uit, afhankelijk van de geselecteerde analyseworkflow.

Analysemethoden

DRAGEN for Illumina DNA Prep with Enrichment Dx voert demultiplexing, FASTQ-bestandengeneratie, bepalingstoewijzing en uitlijning naar een referentiegenoom uit, afhankelijk van de geselecteerde workflow:

- FASTQ-bestanden genereren
- Germline FASTQ- en VCF-bestanden genereren
- Somatic FASTQ- en VCF-bestanden genereren

FASTQ-bestanden genereren

De samengestelde sequenties worden per monster naar FASTQ-bestanden geschreven. FASTQ-bestanden zijn tekstbestanden die sequencinggegevens en kwaliteitsscores voor slechts één monster bevatten. Voor elk monster worden afzonderlijke FASTQ-bestanden gegenereerd per stroomcelbaan, per sequencingbepaling. De naam van het monster zoals gespecificeerd tijdens het instellen van de run wordt opgenomen in de FASTQ-bestandsnaam. FASTQ-bestanden zijn de primaire invoer voor uitlijning. De eerste stap van het genereren van FASTQ-bestanden is demultiplexen. Bij demultiplexen worden clusters die door een filter worden doorgelaten aan een monster toegewezen door elke indexbepalingssequentie te vergelijken met de indexsequenties die zijn opgegeven voor de run. In deze stap wordt geen rekening gehouden met kwaliteitswaarden. Indexbepalingen worden geïdentificeerd aan de hand van de volgende stappen:

- Monsters worden genummerd vanaf 1 op basis van de volgorde waarin ze in de run zijn opgenomen.
- Monster nummer 0 is gereserveerd voor clusters die niet aan een monster zijn toegewezen.
- Clusters worden aan een monster toegewezen wanneer de indexsequentie exact overeenkomt of wanneer er maximaal één discrepantie per indexbepaling is.

De software bevat ORA-compressie om FASTQ-bestanden te comprimeren. Bij gebruik van de ORA-bestandsindeling (*.ora) wordt de md5-controlesom van de FASTQ-inhoud na een compressie- en decompressiecyclus bewaard, om een compressie zonder verlies te garanderen.

DNA-mapping en uitlijning

In de eerste fase van de mapping worden uit de bepaling zaadgetallen gegenereerd en vervolgens wordt gezocht naar exacte overeenkomsten in het referentiegenoom. Deze resultaten worden vervolgens verfijnd door volledige Smith-Waterman-uitlijningen uit te voeren op de locaties met de hoogste dichtheid van overeenkomende zaadgetallen. Dit goed gedocumenteerde algoritme werkt door elke positie van de bepaling te vergelijken met alle kandidaatposities van de referentie. Deze vergelijkingen komen overeen met een matrix van mogelijke uitlijningen tussen de bepaling en de referentie. Voor elk van deze kandidaat-uitlijningsposities genereert Smith-Waterman-scores, die worden gebruikt om te evalueren of de beste uitlijning die door die matrixcel loopt deze bereikt door een nucleotidematch of -mismatch (diagonale beweging), een deletie (horizontale beweging) of een insertie (verticale beweging). Een match tussen bepaling en referentie levert een bonus op de score op en een mismatch of indel brengt een sanctie met zich mee. Het over het geheel genomen hoogst scorende pad door de matrix wordt de gekozen uitlijning.

De specifieke waarden die voor de scores in dit algoritme worden gekozen, geven aan hoe, voor een uitlijning met meerdere mogelijke interpretaties, de mogelijkheid van een indel in tegenstelling tot een of meer SNP's, of de voorkeur voor een uitlijning zonder begrenzing tegen elkaar moeten worden afgewogen. De standaard-DRAGEN-scorewaarden zijn redelijk voor het uitlijnen van bepalingen van gemiddelde lengte op een volledig menselijk referentiegenoom voor toepassingen met variantbepalingen. Elke reeks Smith-Waterman-scoringsparameters vertegenwoordigt een onnauwkeurig model van genomische mutatie- en sequencingfouten. Voor sommige toepassingen kunnen verschillend afgestemde scorewaarden voor uitlijning geschikter zijn.

DRAGEN Germline-variantbepaling

De DRAGEN Germline Small Variant Caller neemt in kaart gebrachte en uitgelijnde DNA-bepalingen als invoer en bepaalt SNP's en indels door een combinatie van kolomgewijze detectie en lokale *de novo* samenstelling van haplotypes.

Eerst worden bepaalbare referentiegebieden geïdentificeerd die voldoende uitlijningsdekking hebben. Binnen deze referentiegebieden identificeert een snelle scan van de gesorteerde bepalingen actieve gebieden, die zijn gecentreerd rond stapelkolommen met aanwijzingen voor een variant. De actieve gebieden zijn opgevuld met voldoende context om belangrijke, niet-referentie-inhoud in de buurt te dekken. Als er aanwijzingen zijn voor indels, worden de actieve gebieden extra opgevuld.

Uitgelijnde bepalingen worden binnen elk actief gebied geknipt en samengevoegd tot een De Bruijn-grafiek. De randen van de geselecteerde bepalingen worden gewogen door het aantal waarnemingen, met de referentiesequentie als ruggengraat. Na enige opschoning en vereenvoudiging van de grafiek worden alle paden van bron naar ontvanger geëxtraheerd als kandidaat-haplotypes. Elk haplotype is via Smith-Waterman uitgelijnd met het referentiegenoom om de varianten die het vertegenwoordigt te identificeren. Deze reeks gebeurtenissen kan worden aangevuld met een detectie op basis van positie.

Voor elk bepaling-haplotypepaar wordt de waarschijnlijkheid $P(r|H)$ dat de bepaling wordt waargenomen, ervan uitgaande dat het haplotype het echte beginmonster is, geschat met behulp van een paarverborgen Markov-model (HMM).

Door per referentiepositie over het actieve gebied te scannen, worden kandidaat-genotypes gevormd uit diploïde combinaties van variantgebeurtenissen (SNP's of indels). Voor elke gebeurtenis (inclusief referentie) wordt de voorwaardelijke kans $P(r|e)$ op het waarnemen van elke overlappende bepaling geschat als de maximale $P(r|H)$ voor haplotypes die de gebeurtenis ondersteunen. Deze worden gecombineerd tot de voorwaardelijke kans $P(r|e_1e_2)$ voor een genotype (gebeurtenispaar) en vermenigvuldigd tot de voorwaardelijke kans $P(R|e_1e_2)$ van het waarnemen van de hele bepalingcombinatie. Met behulp van de formule van Bayes wordt de posterieure kans $P(e_1e_2|R)$ van elk diploïde genotype berekend en wordt de winnaar bepaald.

In de gVCF-modus, die wordt gebruikt voor schaalbare variantbepaling met meerdere monsters, kan de DRAGEN Germline Small Variant Caller per monster worden uitgevoerd om een tussenliggend genomisch variantbepalingsbestand (gVCF) te genereren. De gVCF kan vervolgens worden gebruikt voor de efficiënte gezamenlijke genotypering van meerdere monsters, waardoor de snelle incrementele verwerking van monsters en de opschaling naar grote cohorten mogelijk wordt.

Omdat de DRAGEN Germline Small Variant Caller algoritmen heeft waarmee deze op een efficiënte manier gecorreleerde fouten van echte varianten kan onderscheiden, zijn de filterregels heel eenvoudig.

DRAGEN Somatic-variantbepaling

De DRAGEN Somatic Small Variant Caller neemt in kaart gebrachte en uitgelijnde DNA-bepalingen als invoer en bepaalt SNP's en indels door een lokale *de novo* samenstelling van haplotypes in een actief gebied.

Eerst worden bepaalbare referentiegebieden geïdentificeerd die voldoende uitlijningsdekking hebben. Binnen deze referentiegebieden identificeert een scan van de gesorteerde bepalingen actieve gebieden, die zijn gecentreerd rond stapelkolommen met aanwijzingen voor een variant in de tumorbepalingen. De actieve gebieden zijn opgevuld met voldoende context om belangrijke, niet-referentie-inhoud in de buurt te dekken. Als er aanwijzingen zijn voor indels, krijgen de actieve gebieden extra opvulling.

Uitgelijnde bepalingen worden binnen elk actief gebied geknipt en samengevoegd tot een De Bruijn-grafiek. De randen van de geselecteerde bepalingen worden gewogen door het aantal waarnemingen, met de referentiesequentie als ruggengraat. Na enige opschoning en vereenvoudiging van de grafiek worden alle paden van bron naar ontvanger geëxtraheerd als kandidaat-haplotypes. Elk haplotype is via Smith-Waterman uitgelijnd met het referentiegenoom om de varianten die het vertegenwoordigt te identificeren. Voor elk bepaling-haplotypepaar wordt de waarschijnlijkheid $P(r|H)$ dat de bepaling wordt waargenomen geschat met behulp van een paarverborgen Markov-model (HMM), ervan uitgaande dat het haplotype het echte beginmonster is.

Om de TLOD-score te bepalen, scant de DRAGEN Somatic Small Variant Caller eerst per referentiepositie voor elke kandidaat-somatische gebeurtenis en ook de referentiegebeurtenis over het actieve gebied. De voorwaardelijke kans $P(r|e)$ op het waarnemen van elke overlappende bepaling wordt geschat als de maximale $P(r|H)$ voor haplotypen die de gebeurtenis ondersteunen. Deze worden gecombineerd tot de voorwaardelijke kans $P(r|E)$ voor een gebeurtenishypothese, E , met een mengsel van het referentie- en kandidaat-somatische allel over een reeks van mogelijke allelfrequenties, en vermenigvuldigd om de voorwaardelijke kans $P(R|E)$ te verkrijgen dat de hele bepalingcombinatie wordt waargenomen. Van daaruit wordt een TLOD-score berekend als het bewijs dat een ALT-allel op een bepaalde locus aanwezig is in het tumormonster.

Een run aanmaken

Gebruik de volgende stappen om een run in te stellen in Illumina Run Manager hetzij op de NovaSeq 6000Dx of met behulp van een browser op een netwerkcomputer. Monstergegevens kunnen handmatig worden ingevoerd of door een monsterblad te importeren.

Toepassings- en runinstellingen

1. Selecteer vanuit het scherm Runs **Create Run (Run maken)**.
2. Selecteer de DRAGEN for Illumina DNA Prep with Enrichment Dx-toepassing en selecteer dan **Next (Volgende)**.
3. Voer in het scherm Run Settings (Runinstellingen) een runnaam in. De naam van de run identificeert de run vanaf de sequencing tot en met de analyse.
4. **[Optioneel]** Voer een runbeschrijving in om de run verder te identificeren.
5. Zorg ervoor dat de geselecteerde bibliotheekvoorbereidingskit een bibliotheekvoorbereidingskit voor Illumina DNA Prep with Enrichment Dx is.
6. Selecteer de gewenste indexadapterkit.
7. Voer de bepalinglengte in.
Bepaling 1 en Bepaling 2 hebben een standaardwaarde van 151 cycli.
Index 1 en index 2 hebben een vaste waarde van 10 cycli.
8. **[Optioneel]** Voer een bibliotheekbuisje-ID in.
9. Selecteer **Next (Volgende)**.

Monstergegevens

Gebruik de tabel in het scherm Sample Data (Monstergegevens) om de monsterinformatie handmatig in te voeren. U kunt ook **Import Samples (Monsters importeren)** selecteren om monsterinformatie te uploaden. Raadpleeg [Monsters importeren op pagina 6](#) voor informatie over het importeren van monstergegevens.

Monsters handmatig invoeren

1. Voer in het veld Sample ID (Monster-ID) een unieke identificatie voor het monster in.
2. Gebruik de **Plate - Well Position (Plaat - wellpositie)** om de positie van de well te selecteren. De velden i7 Index, Index 1, i5 Index en Index 2 worden automatisch ingevuld.
3. **[Optioneel]** Voer een bibliotheeknaam in.
4. Voeg rijen toe en herhaal de stappen 1-3 desgewenst totdat alle monsters aan de tabel zijn toegevoegd.
5. Selecteer **Next (Volgende)**.

Monsters importeren

Wanneer u een run plant in Illumina Run Manager met behulp van een browser op een netwerkcomputer, kunt u een sjabloon (*.csv) downloaden in het scherm Sample Data (Monstergegevens).

1. Selecteer **Download Template (Sjabloon downloaden)** om een leeg csv-bestand te downloaden.
2. Voer in het csv-bestand de monstergegevens in en sla het bestand op.
Het csv-bestand van het monsterblad bevat de volgende gegevenskolommen: Monster-ID, Plaats - wellpositie, **Optioneel** Bibliotheeknaam.
3. Selecteer **Import Samples (Monsters importeren)** en blader naar de locatie van het csv-bestand.
4. Selecteer **Next (Volgende)**.

Analyse-instellingen

1. Selecteer de gewenste analyseworkflow:
 - FASTQ-bestanden genereren
 - Germline FASTQ- en VCF-generatie voor een kiemlijnworkflow
 - Somatic FASTQ- en VCF-generatie voor een somatische workflow
2. **[Optioneel]** Selecteer desgewenst het vakje **Generate ORA compressed FASTQs (ORA gecompriëerde FASTQ's genereren)** om FASTQ ORA-compressie in te schakelen.
3. **[Workflows voor het genereren van VCF-bestanden]** Gebruik het keuzemenu **Manifest File Selection (Manifestbestand selecteren)** om een manifestbestand te selecteren.
Een manifestbestand is verplichte invoer voor de DRAGEN for Illumina DNA Prep with Enrichment Dx. Het manifest is een door tabs gescheiden BED-bestand (*.bed), waarin de namen en locaties van de beoogde referentiegebieden worden gespecificeerd.
4. **[Workflow voor het genereren van Somatic FASTQ- en VCF-bestanden]** Gebruik het keuzemenu **Noise File Selection (Ruisbestand selecteren)** om een ruisbestand te selecteren.
Voor het uitfilteren van systematische ruis kan een BED-bestand met een locatiespecifiek ruisniveau worden gespecificeerd. Raadpleeg voor meer informatie [Ruisfiltering op pagina 7](#).
5. Selecteer **Next (Volgende)**.

Run Controleren

1. Bekijk op het scherm Review (Controleren) de informatie die is ingevoerd in de schermen Run Settings (Runinstellingen), Sample Data (Monstergegevens) en Analysis Settings (Analyse-instellingen).
2. Selecteer **Save (Opslaan)**.
De run wordt opgeslagen op het tabblad Planned (Gepland) in het scherm Runs.

Instellingen

Selecteer de app in het scherm Applications (Toepassingen) om de huidige instellingen te bekijken en ze te wijzigen.

Configuratie

Het configuratiescherm toont de volgende toepassingsinstellingen:

- **Library Prep Kits:** Geeft de standaard bibliotheekvoorbereidingskit voor de app weer. Deze instelling kan niet worden gewijzigd.
- **Index Adapter Kits:** Geeft de standaard indexadapterkit voor de app weer. Deze instelling kan niet worden gewijzigd.
- **Bepalingslengtes**— Bepalingslengtes zijn standaard ingesteld op 151 voor de app, maar kunnen tijdens het aanmaken van de run worden gewijzigd.
- **Manifest- en ruisbestanden**— Uploaden en wijzigen van instellingen voor manifest- en ruisbestanden.
 - Selecteer **Upload File (Bestand uploaden)** om bestanden te uploaden voor gebruik bij analyses.
 - Selecteer het keuzerondje **Default (Standaard)** om het bestand als het standaardmanifestbestand of standaardruisbestand in te stellen dat tijdens het maken van de run wordt geselecteerd wanneer de toepassing is geselecteerd.
 - Schakel het selectievakje **Enabled (Ingeschakeld)** in om in te stellen dat het bestand tijdens het maken van de run in het vervolkeuzemenu wordt weergegeven.

Rechten

Gebruik de selectievakjes op het scherm Permissions (Machtigingen) om de gebruikerstoegang voor de app te beheren.

Ruisfiltering

Bij gebruik van de somatische workflow is systematische ruisfiltering beschikbaar. Het filter kan worden gebruikt in de modus tumor-normaal, maar is vooral nuttig voor alleen-tumor-runs waarbij een overeenkomende normaal niet beschikbaar is.

De systematische ruis-BED moet worden gegenereerd op basis van normale monsters. Het wordt aanbevolen om systematische ruisbestanden te bouwen die bibliotheekpreparatie-, sequencingsysteem- en panelspecifiek zijn. Voor het genereren van ruisbestanden wordt aanbevolen om ongeveer 50 normale monsters te gebruiken.

Analyse-uitvoerbestanden

DRAGEN for Illumina DNA Prep with Enrichment Dx slaat de volgende informatie op in de analysemap. Alleen de kiemlijn- en somatische workflows produceren een PDF.

- Gebruikt manifestbestand
- Softwareversies
- Monster-ID's
- Totaal aantal uitgelijnde bepalingen
- Percentage uitgelijnde bepalingen per monster
- Aantal bepaalde SNV's per monster
- Aantal bepaalde indels per monster
- Dekkingsstatistieken

Uitvoerbestanden van analyses

De toepassing genereert de volgende uitvoerbestanden. Welke bestanden precies worden gegenereerd, hangt af van de gebruikte analyseworkflow. De uitvoerbestanden bevinden zich in de analysemap.

| Uitvoerbestand | Omschrijving |
|-------------------------------------|---|
| FASTQ (*.fastq.gz of *.fastq.ora) | Tussentijdse bestanden die basebepalingen met kwaliteitsscores bevatten. FASTQ-bestanden zijn de primaire invoer voor de uitlijningsstap. Als ORA-compressie geselecteerd is, geeft de bestandsnaam dat weer. |
| Uitlijning BAM-bestanden (*.bam) | Bevat uitgelijnde bepalingen voor een bepaald monster. |
| Genoom VCF-bestanden (*.gvcf.gz) | Bevat het genotype voor elke positie, of het nu als variant of als referentie is bepaald. |
| VCF-bestanden (*.vcf.gz) | Bevat varianten die op elke positie worden bepaald. |
| Rapport met runstatistieken (*.csv) | Bevat kwaliteitsstatistieken over de run, waaronder de totale opbrengst en de Q30-score. |

FASTQ-bestanden

FASTQ (*.fastq.gz, *.fastq.ora) is een tekstbestandsindeling die basebepalingen en kwaliteitswaarden per bepaling bevat. Elk bestand bevat de volgende informatie:

- De monsteridentificatie
- De sequentie
- Een plusteken (+)
- De Phred-kwaliteitsscores in een ASCII + 33-indeling

De monsteridentificatie heeft de volgende indeling.

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y
ReadNum:FilterFlag:0:SampleNumber
Voorbeeld:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAAA9#:<#<;<<<????#=#
```

BAM-bestanden

Een BAM-bestand (*.bam) is de gecomprimeerde binaire versie van een SAM-bestand (Sequence Alignment Map) dat wordt gebruikt om uitgelijnde sequenties tot 128 Mb weer te geven. BAM-bestanden hebben de naam `SampleName_S#.bam`. Hierbij is # het monsternummer dat wordt bepaald door de plaats van het monster in de runlijst. In de multinode-modus wordt de S# ingesteld op S1, ongeacht de volgorde van het monster.

BAM-bestanden bevatten een kopregel- en een uitlijningssectie:

- Kopregel: bevat informatie over het gehele bestand, zoals de monsternaam, monsterlengte en uitlijningsmethode. Uitlijningen in de uitlijningssectie zijn gekoppeld aan specifieke informatie in de kopregelsectie.
- Uitlijningen: bevat de bepalingnaam, bepalingsvolgorde, bepalingskwaliteit, uitlijningsinformatie en aangepaste tags. De bepalingnaam bevat het chromosoom, de startcoördinaat, de uitlijningskwaliteit en de tekenreeks met overeenkomende beschrijvende elementen.

Het uitlijningsgedeelte bevat de volgende informatie voor elke bepaling of bepalingspaar:

- AS: Kwaliteit van paired-end-uitlijning.
- RG: Bepalingsgroep, die het aantal bepalingen voor een specifiek monster aangeeft.
- BC: Barcodetag, die de gedemultiplexte monster-ID aangeeft die bij de bepalingswaarde hoort.
- SM: Kwaliteit van enkele einde-uitlijning.

- XC: Tekenreeks met overeenkomende beschrijvende elementen.
- XN: Amplicon-naamtag, die de amplicon-ID registreert die bij de BAM-indexbestanden (*.bam.bai) van de bepaling hoort geeft een index van het corresponderende BAM-bestand.

VCF-bestanden

Variantbepalingsbestanden (*.vcf) bevatten informatie over varianten die op specifieke posities in een referentiegenoom zijn gevonden.

De kopregel van het VCF-bestand bevat de versie van de VCF-bestandsindeling en de versie van de variantbepaler en vermeldt de annotaties die in de rest van het bestand worden gebruikt. De VCF-kopregel vermeldt ook het referentiegenoombestand en het BAM-bestand. De laatste regel in de kopregel bevat de kolomtitels voor de gegevensregels. Alle gegevensregels van het VCF-bestand bevatten informatie over één variant.

Tabel 1 VCF-bestandskopregels

| Kopregel | Omschrijving |
|----------|--|
| CHROM | Het chromosoom van het referentiegenoom. Chromosomen verschijnen in dezelfde volgorde als het FASTA-referentiebestand. |
| POS | De enkele basepositie van de variant in het referentiechromosoom. Voor enkelvoudige nucleotidevarianten (SNV's) is deze positie de referentiebasis bij de variant. Voor indels is deze positie de referentiebasis die onmiddellijk aan de variant voorafgaat. |
| Id | Het rs-nummer (referentie-SNP) voor de SNP verkregen uit <code>dbSNP.txt</code> , indien van toepassing. Als er meerdere rs-nummers op deze locatie bestaan, wordt de lijst door puntkomma's gescheiden. Als op deze positie geen dbSNP-gegeven bestaat, wordt een markering voor ontbrekende waarde ('.') gebruikt. |
| REF | Het referentiegenotype. Zo wordt een verwijdering van één T weergegeven als referentie-TT en alternatieve T. Een A-naar-T-variant van één nucleotide wordt weergegeven als referentie A en alternatieve T. |
| ALT | De allelen die verschillen van de referentiebepaling. Bijvoorbeeld, een invoeging van een enkele T wordt weergegeven als referentie A en alternatieve AT. Een A-naar-T-variant van één nucleotide wordt weergegeven als referentie A en alternatieve T. |

| Kopregel | Omschrijving |
|----------|--|
| QUAL | Een kwaliteitsscore op de Phred-schaal die is toegekend door de variantbepaler. Hogere scores wijzen op een groter vertrouwen in de variant en een lagere waarschijnlijkheid van aanwezige fouten. Voor een kwaliteitsscore van Q is de geschatte kans op een fout $10^{-(Q/10)}$. De reeks Q30-bepalingen heeft bijvoorbeeld een foutenpercentage van 0,1%. Veel variantbepalers kennen kwaliteitsscores toe op basis van hun statistische modellen, die hoog zijn in verhouding tot het waargenomen foutenpercentage. |

Tabel 2 VCF-bestandsannotaties

| Kopregel | Omschrijving |
|--------------------|---|
| FILTER (Filter) | <p>Als aan alle filters is voldaan, wordt PASS in de filterkolom geschreven.</p> <p>Mogelijke FILTER-vermeldingen voor de Germline-workflow zijn onder meer:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL: Toegepast als de QUAL-score van de SNP-variant niet aan de drempelwaarde voldoet • DRAGENIndelHardQUAL: Toegepast als de QUAL-score van indelvariant niet aan de drempelwaarde voldoet • LowDepth: Locatie gefilterd omdat de dekkingsdiepte niet voldoet aan de drempelwaarde • LowGQ: Locatie gefilterd omdat de kwaliteit van het genotype niet voldoet aan de drempelwaarde • PloidyConflict: Genotypebepaling van variantbepaling niet consistent met chromosoomploïdie • base_quality: Locatie gefilterd omdat de mediane basekwaliteit van alt-bepalingen op deze locus niet voldoet aan de drempelwaarde • filtered_reads: Locatie gefilterd omdat een te groot deel van de bepalingen is uitgefilterd • fragment_length: Locatie gefilterd omdat het absolute verschil tussen de mediane fragmentlengte van alt-bepalingen en de mediane fragmentlengte van ref-bepalingen op deze locus de drempelwaarde overschrijdt • low_depth: Locatie gefilterd omdat de bepalingdiepte te laag is • low_frac_info_reads: Locatie gefilterd omdat het deel van informatieve bepalingen onder de drempelwaarde ligt • low_normal_depth: Locatie gefilterd omdat de normale bepalingdiepte van het monster te laag is • long_indel: Locatie gefilterd omdat de indellengte te lang is • mapping_quality: Locatie gefilterd omdat de mediane mappingkwaliteit van alt-bepalingen op deze locus niet voldoet aan de drempelwaarde • multiallelic: Locatie gefilterd omdat meer dan twee alt-allelen door de tumor-LOD komen • non_homref_normal: Locatie gefilterd omdat het normale monstergenotype geen homozygote referentie is • no_reliable_supporting_read: Locatie gefilterd omdat er geen betrouwbare ondersteunende somatische bepaling bestaat • panel_of_normals: Gezien in ten minste één monster in het panel van normalen-vcf • read_position: Locatie gefilterd omdat de mediaan van de afstanden tussen begin/einde van de bepaling en deze locus onder de drempelwaarde ligt • RMxNRepeatRegion: Locatie gefilterd omdat het variantallel geheel of gedeeltelijk een herhaling van de referentie is • strand_artifact: Locatie gefilterd vanwege ernstige strengvertekening • str_contraction: Locatie gefilterd als gevolg van een vermoedelijke PCR-fout waarbij het alt-allel één herhalingseenheid minder is dan de referentie |

| Kopregel | Omschrijving |
|---------------------|--|
| FILTER (vervolg) | <ul style="list-style-type: none"> • too_few_supporting_reads: Locatie gefilterd omdat er te weinig ondersteunende bepalingen in het tumormonster zijn • weak_evidence: Somatische variantscore voldoet niet aan de drempelwaarde <p>Mogelijke FILTER-vermeldingen voor de somatische workflow zijn onder meer:</p> <ul style="list-style-type: none"> • base_quality: Locatie gefilterd omdat de mediane basekwaliteit van alt-bepalingen op deze locus niet voldoet aan de drempelwaarde • filtered_reads: Locatie gefilterd omdat een te groot deel van de bepalingen is uitgefilterd • fragment_length: Locatie gefilterd omdat het absolute verschil tussen de mediane fragmentlengte van alt-bepalingen en de mediane fragmentlengte van ref-bepalingen op deze locus de drempelwaarde overschrijdt • low_depth: Locatie gefilterd omdat de bepalingstiepte te laag is • low_frac_info_reads: Locatie gefilterd omdat het deel van informatieve bepalingen onder de drempelwaarde ligt • low_normal_depth: Locatie gefilterd omdat de normale bepalingstiepte van het monster te laag is • long_indel: Locatie gefilterd omdat de indellengte te lang is • mapping_quality: Locatie gefilterd omdat de mediane mappingkwaliteit van alt-bepalingen op deze locus niet voldoet aan de drempelwaarde • multiallelic: Locatie gefilterd omdat meer dan twee alt-allelen door de tumor-LOD komen • non_homref_normal: Locatie gefilterd omdat het normale monstergenotype geen homozygote referentie is • no_reliable_supporting_read: Locatie gefilterd omdat er geen betrouwbare ondersteunende somatische bepaling bestaat • panel_of_normals: Gezien in ten minste één monster in het panel van normalen-vcf • read_position: Locatie gefilterd omdat de mediaan van de afstanden tussen begin/einde van de bepaling en deze locus onder de drempelwaarde ligt • RMxNRepeatRegion: Locatie gefilterd omdat het variantallel geheel of gedeeltelijk een herhaling van de referentie is • strand_artifact: Locatie gefilterd vanwege ernstige strengvertekening • str_contraction: Locatie gefilterd als gevolg van een vermoedelijke PCR-fout waarbij het alt-allel één herhalingseenheid minder is dan de referentie • too_few_supporting_reads: Locatie gefilterd omdat er te weinig ondersteunende bepalingen in het tumormonster zijn • weak_evidence: Somatische variantscore voldoet niet aan de drempelwaarde • systematic_noise: Locatie gefilterd op basis van bewijs van systematische ruis in normalen |

| Kopregel | Omschrijving |
|----------------------|---|
| INFO (Informatie) | <p>Mogelijke INFO-vermeldingen voor de Germline-workflow zijn onder meer:</p> <ul style="list-style-type: none"> • AC: Aantal allelen in genotypen voor elk ALT-allel, in dezelfde volgorde als vermeld. • AF: Allelfrequentie voor elk ALT-allel, in dezelfde volgorde als vermeld. • AN: Het totale aantal allelen in de bepaalde genotypen. • DB: dbSNP-lidmaatschap. • FS: Een p-waarde op de Phred-schaal met Fisher's exacte test om vertekening van de streng op te sporen. • QD: Variantvertrouwen/kwaliteit per diepte. • R2_5P_bias: Score gebaseerd op partnerbias en afstand tot 5 prime end. • SOR: Symmetric Odds Ratio van 2x2 contingentietabel om strengvertekening te detecteren. • DP: Geschatte bepalingdiepte (informatief en niet-informatief); sommige bepalingen kunnen zijn gefilterd op basis van mapq enz. • END: Stoppositie van het interval. • FractionInformativeReads: De fractie informatieve bepalingen op het totaal aantal bepalingen. • MQ: RMS-mappingkwaliteit. • MQRankSum: Z-score van Wilcoxon-ranksomtest van Alt vs. Ref bepalingmappingkwaliteiten. • ReadPosRankSum: Z-score van Wilcoxon-ranksomtest van Alt vs. Ref bepalingpositievertekening. • SOMATIC: Ten minste één variant op deze positie is somatisch. <p>Mogelijke INFO-vermeldingen voor de somatische workflow zijn onder meer:</p> <ul style="list-style-type: none"> • DP: Geschatte bepalingdiepte (informatief en niet-informatief); sommige bepalingen kunnen zijn gefilterd op basis van mapq enz. • END: Stoppositie van het interval. • FractionInformativeReads: De fractie informatieve bepalingen op het totaal aantal bepalingen. • MQ: RMS-mappingkwaliteit. • MQRankSum: Z-score van Wilcoxon-ranksomtest van Alt vs. Ref bepalingmappingkwaliteiten. • ReadPosRankSum: Z-score van Wilcoxon-ranksomtest van Alt vs. Ref bepalingpositievertekening. • AQ: Systematische geluidsscore. • hotspot: Bekende somatische locatie, gebruikt om het vertrouwen in de bepaling te vergroten. • SOMATIC: Ten minste één variant op deze positie is somatisch. |

| Kopregel | Omschrijving |
|----------------------|--|
| FORMAT (Indeling) | <p>In de indelingskolom staan de velden gescheiden door dubbele punten. Bijvoorbeeld: GT:GQ.</p> <p>De beschikbare velden voor de Germline-workflow zijn:</p> <ul style="list-style-type: none"> • AD: Allelische diepten (alleen informatieve bepalingen tellen uit het totale aantal bepalingen) voor de ref- en alt-allelen in de vermelde volgorde. • AF: Allelfracties voor alt-allelen in de aangegeven volgorde. • DP: Geschatte bepalingdiepte (bepalingen met MQ=255 of met slechte paren worden gefilterd). • F1R2: Aantal bepalingen in F1R2-paaroriëntatie die elk allel ondersteunen. • F2R1: Aantal bepalingen in F2R1-paaroriëntatie die elk allel ondersteunen. • GP: Posterieure kansen op de Phred-schaal voor genotypen zoals gedefinieerd in de VCF-specificatie. • GQ: Kwaliteit van het genotype. • GT: Genotype. 0 komt overeen met de referentiebase, 1 komt overeen met de eerste vermelding in de kolom ALT, enz. De schuine streep (/) geeft aan dat er geen faseringsinformatie beschikbaar is. • MB: Componentstatistieken per monster ter detectie van partnervetekening. • PL: Genormaliseerde waarschijnlijkheden op de Phred-schaal voor genotypes zoals gedefinieerd in de VCF-specificatie. • PRI: Eerdere kansen op de Phred-schaal voor genotypen. • PS: ID-informatie over fysieke faserings, waarbij elke unieke ID binnen een bepaald monster (maar niet tussen monsters) records binnen een faseringsgroep verbindt. • SB: Componentstatistieken per monster waaruit de Fisher's Exact Test is samengesteld ter detectie van strengvertekening. • SQ: Somatische kwaliteit. <p>De beschikbare velden voor de somatische workflow zijn:</p> <ul style="list-style-type: none"> • AD: Allelische diepten (alleen informatieve bepalingen tellen uit het totale aantal bepalingen) voor de ref- en alt-allelen in de vermelde volgorde. • AF: Allelfracties voor alt-allelen in de aangegeven volgorde. • DP: Geschatte bepalingdiepte (bepalingen met MQ=255 of met slechte paren worden gefilterd). • F1R2: Aantal bepalingen in F1R2-paaroriëntatie die elk allel ondersteunen. • F2R1: Aantal bepalingen in F2R1-paaroriëntatie die elk allel ondersteunen. • GT: Genotype. 0 komt overeen met de referentiebase, 1 komt overeen met de eerste vermelding in de kolom ALT, enz. De schuine streep (/) geeft aan dat er geen faseringsinformatie beschikbaar is. • MB: Componentstatistieken per monster ter detectie van partnervetekening. • PS: ID-informatie over fysieke faserings, waarbij elke unieke ID binnen een bepaald monster (maar niet tussen monsters) records binnen een faseringsgroep verbindt. • SB: Componentstatistieken per monster waaruit de Fisher's Exact Test is samengesteld ter detectie van strengvertekening. |

| Kopregel | Omschrijving |
|----------------------|---|
| FORMAT (Indeling) | <ul style="list-style-type: none">• SQ: Somatische kwaliteit. |
| SAMPLE (Monster) | In de kolom Sample (Monster) staan de waarden die in de kolom FORMAT (Indeling) zijn opgegeven. |

Genome VCF-bestanden

Genome VCF-bestanden (*.gvcf.gz) volgen een reeks conventies om alle locaties binnen het genoom in een redelijk compact formaat weer te geven. De gVCF-bestanden bevatten alle locaties binnen het gebied van belang in één bestand voor elk monster. Het gVCF-bestand toont niet-bepalingen op posities die niet door alle filters komen. De genotypetag (GT) ./. geeft een niet-bepaling aan.

Analyseresultaten weergeven

Lopende runs worden weergegeven in het tabblad Active (Actief). Voltooides runs worden weergegeven in het tabblad Completed (Voltooid). Raadpleeg de [NovaSeq 6000Dx-productdocumentatie \(documentnr. 200010105\)](#) voor meer informatie over het bekijken van de resultaten.

Technische ondersteuning

Voor technische ondersteuning neemt u contact op met Illumina Technische ondersteuning.

Website: www.illumina.com
E-mail: techsupport@illumina.com

Telefoonnummers voor technische ondersteuning van Illumina

| Regio | Gratis telefoonnummer | Internationaal |
|-----------------|-----------------------|-------------------|
| Australië | +61 1800 775 688 | |
| Oostenrijk | +43 800 006249 | +43 1 9286540 |
| België | +32 800 77 160 | +32 3 400 29 73 |
| Canada | +1 800 809 4566 | |
| China | | +86 400 066 5835 |
| Denemarken | +45 80 82 01 83 | +45 89 87 11 56 |
| Finland | +358 800 918 363 | +358 9 7479 0110 |
| Frankrijk | +33 8 05 10 21 93 | +33 1 70 77 04 46 |
| Duitsland | +49 800 101 4940 | +49 89 3803 5677 |
| Hongkong, China | +852 800 960 230 | |
| India | +91 8006500375 | |
| Indonesië | | 0078036510048 |
| Ierland | +353 1800 936608 | +353 1 695 0506 |
| Italië | +39 800 985513 | +39 236003759 |
| Japan | +81 0800 111 5011 | |
| Maleisië | +60 1800 80 6789 | |
| Nederland | +31 800 022 2493 | +31 20 713 2960 |
| Nieuw-Zeeland | +64 800 451 650 | |
| Noorwegen | +47 800 16 836 | +47 21 93 96 93 |
| Filippijnen | +63 180016510798 | |
| Singapore | 1 800 5792 745 | |
| Zuid-Korea | +82 80 234 5300 | |
| Spanje | +34 800 300 143 | +34 911 899 417 |

| Regio | Gratis telefoonnummer | Internationaal |
|----------------------------|------------------------------|-----------------------|
| Zweden | +46 2 00883979 | +46 8 50619671 |
| Zwitserland | +41 800 200 442 | +41 56 580 00 00 |
| Taiwan, China | +886 8 06651752 | |
| Thailand | +66 1800 011 304 | |
| Verenigd Koninkrijk | +44 800 012 6019 | +44 20 7305 7197 |
| Verenigde Staten | +1 800 809 4566 | +1 858 202 4566 |
| Vietnam | +84 1206 5263 | |

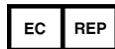
Veiligheidsinformatiebladen – zijn verkrijgbaar op de website van Illumina via support.illumina.com/sds.html.

Productdocumentatie – kan in pdf-formaat worden gedownload via support.illumina.com.



Illumina
5200 Illumina Way
San Diego, Californië 92122 VS
+1 800 809 ILMN (4566)
+1 858 202 4566 (buiten Noord-Amerika)
techsupport@illumina.com
www.illumina.com

CE



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

Australische sponsor

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

BESTEMD VOOR IN-VITRODIAGNOSTIEK

© 2022 Illumina, Inc. Alle rechten voorbehouden.

illumina[®]