

DRAGEN for Illumina DNA Prep with Enrichment Dx on NextSeq 550Dx

Gebruikershandleiding voor de toepassing

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).

© 2023 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 informatie over specifieke handelsmerken.

Revisiegeschiedenis

Document	Datum	Omschrijving van wijziging
200025238 v00	Februari 2023	Eerste uitgave.

Inhoudsopgave

Revisiegeschiedenis	iii
Overzicht	1
Analysemethoden	1
Een geplande run maken	5
Instellingen	8
Manifestbestand	8
Ruisfiltering (optioneel)	9
Analyse-uitvoerbestanden	9
FASTQ-bestanden	11
BAM-bestanden	11
VCF-bestanden	12
Analyse opnieuw in de wachtrij plaatsen	19
Technische ondersteuning	21

Overzicht

De DRAGEN for Illumina DNA Prep with Enrichment Dx toepassing (DRAGEN for IDPE Dx) wordt gebruikt voor het plannen en uitvoeren van secundaire analyse van IDPE Dx-bibliotheken die zijn gegenereerd voor sequencing op de NextSeq 550Dx.

DRAGEN for IDPE Dx ondersteunt sequencing voor analyse wanneer gebruikt met de Illumina DNA Prep with Enrichment Dx Library Prep, NextSeq 550Dx, en Illumina DRAGEN Server for NextSeq 550Dx.

Analysemethoden

DRAGEN for IDPE Dx voert demultiplexing, FASTQ-bestandengeneratie, bepalingstoewijzing, uitlijning naar een referentiegenoom en kleine variantbepaling uit, afhankelijk van de geselecteerde workflow:

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

OPMERKING ORA-compressie is beschikbaar voor gebruik met alle drie de workflows. DRAGEN ORA Compression is een volledig verliesloze compressiesoftware die een bestand creëert met de extensie Original Read Archive (*.ora). Het ora-formaat is een referentiegebaseerd compressieformaat voor FASTQ-bestanden en is ontworpen voor zeer snelle compressie/decompressie en een hoge compressieverhouding.

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. Dit formaat kan optioneel worden ingeschakeld. 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

Na het genereren van FASTQ worden bepalingen in kaart gebracht en uitgelijnd met een referentiegenoom. 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. Het algoritme wordt hardware-geaccelereerd op de DRAGEN field-programmeable gate array (FPGA)-kaarten. Het referentiegenoom dat in de app wordt gebruikt, wordt gemaakt vanuit de UCSC hg19 FASTA met de DRAGEN-optie om een op liftover gebaseerde alt-aware hashtable te maken.

DRAGEN kiemlijnvariantbepaling

De DRAGEN Germline Small Variant Caller neemt in kaart gebrachte en uitgelijnde DNA-bepalingen als invoer en bepaalt enkel-nucleotide-polymorfieën en invoeging of verwijdering (indels) door een combinatie van kolomsgewijze detectie en lokale *de novo* samenstelling van haplotypes. Selecteer de workflow van de DRAGEN Germline Small Variant Caller, selecteer de kiemlijnworkflow.

Variantbepaling van de kiemlijn wordt meestal gebruikt voor kiemlijnmonsters waarvan bekend is dat er twee ploëdiën zijn. 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 op 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, 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. 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.

DRAGEN for IDPE Dx past automatisch filteren toe. Raadpleeg [Kiemlijnworkflow VCF-bestandsannotaties op pagina 14](#) voor meer informatie.

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. Selecteer een somatische varianttoepassing om de DRAGEN Somatic Small Variant Caller in te schakelen.

Somatische variantbepaling wordt doorgaans gebruikt voor tumormonsters. Met deze workflow maakt DRAGEN geen aannames over ploïdiën, waardoor detectie van laagfrequente allelen mogelijk is. Voor loci met een dekking tot 100x in het tumormonster heeft DRAGEN een detectiedrempel bij variante allelfrequenties van 5%. De limiet schaalte met toenemende diepte per locus en halveert elke keer dat de dekking verdubbelt tot meer dan 100x. Eerst worden bepaalde referentiegebieden geïdentificeerd die voldoende uitlijningsdekking hebben. Binnen deze referentiegebieden identificeert een scan van de gesorteerde bepalingen actieve gebieden, die zijn gecentreerd op stapelkolommen met aanwijzingen voor een variant in de tumorbevestigingen. 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 score van de detectielimiet voor tumoren (TLOD) te bepalen, scant DRAGEN Somatic Small Variant Caller over het actieve gebied eerst op referentiepositie voor elke kandidaat voor een somatische gebeurtenis en de referentiegebeurtenis. 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 $P(r|E)$ 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.

DRAGEN for IDPE Dx past automatisch filteren toe. Raadpleeg [Somatische workflow VCF-bestandsannotaties op pagina 17](#) voor meer informatie.

Een geplande run maken

Gebruik de volgende stappen om een run in te stellen in Illumina Run Manager hetzij op de NextSeq 550Dx of met behulp van een browser op een netwerkcomputer. Gebruik een browser op een netwerkcomputer als u monstergegevens wilt importeren. Raadpleeg de Softwarehandleiding Illumina Run Manager voor NextSeq 550Dx (documentnr. 200025239) voor instructies over toegang Illumina Run Manager vanaf een netwerkcomputer.

Er zijn twee verschillende manieren om een nieuwe geplande run te maken:

- **Import Run** (Run importeren): Gebruik een monsterblad van een bestaande run als sjabloon voor een nieuwe run. Raadpleeg de Softwarehandleiding Illumina Run Manager voor NextSeq 550Dx (documentnr. 200025239) voor informatie over het importeren van een run.
- **Create Run** (Run maken): Voer handmatig runparameters in. De volgende instructies beschrijven hoe u een handmatige run kunt aanmaken.

OPMERKING De vereiste invoervelden in de gebruikersinterface zijn gemarkeerd met een sterretje (*).

Application (Toepassing)

1. Selecteer van het scherm Runs op het tabblad Planned (Gepland) **Create Run (Run aanmaken)**.
2. Selecteer de DRAGEN for Illumina DNA Prep with Enrichment Dx-toepassing en selecteer dan **Next** (Volgende).

Runinstellingen

1. Voer in het scherm Run Settings (Runinstellingen) een unieke runnaam in. De naam van de run identificeert de run vanaf de sequencing tot en met de analyse.
2. **[Optioneel]** Voer een runbeschrijving in om de run verder te identificeren.
3. Selecteer de indexadapterkit(s) die tijdens de bibliotheekvoorbereiding wordt/worden gebruikt.
4. Controleer de Lengte van de Bepaling en wijzig deze indien nodig. Bepaling 1 en Bepaling 2 hebben een standaardwaarde van 151 cycli. Index 1 en index 2 hebben een vaste waarde van 10 cycli en kunnen niet worden aangepast.
5. **[Optioneel]** Voer een bibliotheekbuisje-ID in.
6. Selecteer **Next** (Volgende).

Monstergegevens

Monstergegevens omvatten de monster-ID, well-positie (well-positie van de indexplaat) en bibliotheeknaam. Bij gebruik van index A&B bevat de well-positie ook plaatidentificatie.

Er zijn twee manieren om monstergegevens in te voeren:

- **Import Samples** (Monsters importeren): Gebruik een sjabloonbestand dat beschikbaar is om te downloaden in het scherm Sample Data (Monstergegevens).
- **Manually** (Handmatig): Voer de monstergegevens rechtstreeks in de tabel in het scherm Sample Data (Monstergegevens) in.

Monsters importeren

Wanneer u een sequencing-run plant met behulp van een browser op een netwerkcomputer, is er een sjabloonbestand (*.csv) beschikbaar om te downloaden in het scherm Sample Data (Monstergegevens). Het sjabloonbestand is niet beschikbaar om te downloaden wanneer u toegang krijgt Illumina Run Manager tot de NextSeq 550Dx besturingssysteemsoftware. Voer de volgende stappen uit om monstergegevens in te voeren met de functie Import Samples (Monsters importeren).

OPMERKING Voer de stappen van Run Settings (Runinstellingen) uit voordat u verdergaat.

1. Selecteer **Download Template** (Sjabloon downloaden) om een leeg csv-bestand te downloaden.
2. Voer vanuit het sjabloonbestand monstergegevens in en sla het bestand vervolgens op. De bibliotheeknaam is optioneel.

OPMERKING Bij gebruik van Index A&B moeten de gegevens voor kolom B zowel de plaat- als de well-positie (well-positie van de indexplaat) bevatten. Voorbeeld: A-A01, A-A02, A-A03.

3. Selecteer **Import Samples** (Monsters importeren) en blader naar het sjabloonbestand met de monstergegevens van de vorige stap.
4. Selecteer **Open** (openen), **Proceed** (Doorgaan) en vervolgens **Next** (Volgende).

OPMERKING Het wijzigen van de monster-ID voordat u Next (Volgende) selecteert, kan leiden tot een fout. Om fouten te voorkomen, voltooi het instellen van de run voordat u wijzigingen aanbrengt.

Monsters handmatig invoeren

Gebruik de tabel in het scherm Sample Data (Monstergegevens) om de monstergegevens handmatig in te voeren.

1. Voer in het veld Sample ID (Monster-ID) een unieke identificatie voor het monster in.
2. Gebruik **Well Position** (well-positie) (index A of index B) of **Plate - Well Position** (plaat - well-positie) (index A&B) om de bijbehorende index voor de monsters 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. U kunt meerdere rijen tegelijk toevoegen door eerst het aantal rijen in te voeren dat u wilt toevoegen en vervolgens het +-pictogram te selecteren. U kunt ook rijen verwijderen door het vakje naast het rijnummer te selecteren en vervolgens op het prullenbakpictogram te klikken.
5. Selecteer **Next** (Volgende).

Analyse-instellingen

1. Selecteer de gewenste analyseworkflow:
 - FASTQ-bestanden genereren
 - FASTQ- en VCF-bestanden genereren voor een kiemlijnworkflow (manifestbestand vereist)
 - FASTQ- en VCF-bestanden genereren voor een somatische workflow (manifestbestand vereist)
2. **[Optioneel] Generate ORA compressed FASTQs** (Genereren van ORA-gecomprimeerde FASTQ's) is standaard ingeschakeld. FASTQ ORA-compressie comprimeert FASTQ-bestanden zonder verlies tot 5x in vergelijking met fastq.gz. Schakel **Generate ORA compressed FASTQs** (ORA-gecomprimeerde FASTQ-bestanden genereren) uit als ongecomprimeerde gegevens (fastq.gz) de voorkeur hebben.
3. Voor kiemlijn- en somatische workflows is een manifestbestand vereist. Gebruik het keuzemenu **Manifest File Selection** (Manifestbestand selecteren) om een manifestbestand te selecteren. Het manifest is een door tabs gescheiden BED-bestand (*.bed), waarin de namen en locaties van de beoogde referentiegebieden worden gespecificeerd. Raadpleeg voor meer informatie [Manifestbestand op pagina 8](#).
4. **[Optioneel]** Gebruik voor somatische workflows het keuzemenu **Noise File Selection** (Ruisbestand selecteren) om een systematisch ruisbestand te selecteren.
Voor het uitfilteren van systematische ruis kan een BED(*.bed.gz)-bestand met een locatiespecifiek ruisniveau worden gespecificeerd. Raadpleeg voor meer informatie [Ruisfiltering \(optioneel\) op pagina 9](#).
5. Selecteer **Next** (Volgende).

Run Controleren

1. Bekijk op het scherm Review (Controleren) de informatie voor 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

Om DRAGEN for IDPE Dx toepassingsinstellingen te bekijken of te wijzigen, selecteert u eerst het pictogram Toepassingen op het hoofdscherm. Selecteer vervolgens de toepassing die u wilt bekijken of wijzigen. Een beheerdersaccount is vereist om instellingen te wijzigen.

Configuratie

Het configuratiescherm toont de volgende toepassingsinstellingen:

- **Library Prep Kits** (Bibliotheekvoorbereidingssets): Geeft de standaard bibliotheekvoorbereidingskit voor de app weer. Deze instelling kan niet worden gewijzigd.
- **Index Adapter Kits** (Indexadapterkits): Geeft de standaard indexadapterkit voor de app weer. Deze instelling kan niet worden gewijzigd.
- **Read lengths** (Bepalingslengtes): Bepalingslengtes zijn standaard ingesteld op 151 voor de app, maar kunnen tijdens het aanmaken van de run worden gewijzigd.
- **Manifest and Noise Files** (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 vervolgkeuzemenu wordt weergegeven.

Rechten

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

Manifestbestand

Bij gebruik van DRAGEN for IDPE Dx is een manifestbestand vereist voor de volgende workflows:

- FASTQ- en VCF-bestanden genereren voor een kiemlijnworkflow
- FASTQ- en VCF-bestanden genereren voor een somatische workflow

Het manifestbestand is een door tabs gescheiden tekstbestand dat gebruik maakt van het BED-formaat (*.bed), waarin de namen en locaties van de beoogde referentiegebieden worden gespecificeerd. Het hoofdgedeelte van het manifestbestand is het gedeelte Regions (Regio's) en moet de volgende gegevenskolommen bevatten:

Kolom	Beschrijving
Naam	Unieke door de gebruiker gespecificeerde naam voor het doel
Chromosoom	Chromosoomlocatie (bijv. chr10, chr5, enz.)
Start	1-gebaseerde index voor de startpositie van het doel
Stop	1-gebaseerde index voor de stoppositie van het doel
Lengte upstreamsonde	De lengte van de upstreamsonde. Voor de DRAGEN for IDPE Dx app moet dit worden ingesteld op 0.
Lengte downstreamsonde	De lengte van de downstreamsonde. Voor de DRAGEN for IDPE Dx app moet dit worden ingesteld op 0.

OPMERKING Een geldig manifestbestandsformaat is vereist voor analyse. DRAGEN stopt de analyse als het manifestbestand ongeldig is.

Ruisfiltering (optioneel)

Het systematische ruisfilter is beschikbaar voor somatische variantbepaling en kan worden gebruikt om vals-positieve bepalingen te verminderen door rekening te houden met locatiespecifieke ruis. Het systematische ruisbestand wordt gegenereerd door eerst ongeveer 50 normale monsters te verzamelen (bij voorkeur specifiek voor het panel, bibliotheekvoorbereiding en sequencer) en vervolgens wordt de som van allelfrequenties onder 30% op elke locatie met voldoende dekking gedeeld door het totale aantal monsters (allelfrequenties boven 30% worden verondersteld kiemlijnvarianten te zijn en geen ruis). Zodra de ruiswaarden zijn gegenereerd, worden somatische varianten die op die locatie worden gedetecteerd, gefilterd.

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. Het systematische ruisbestand moet een BED-bestand gebruiken met de bestandsextensie (*.bed.gz) en moet vier kolommen bevatten: Chromosoom-, start-, eind- en locatiespecifieke geluidsniveaus voor elke rij. Systematische ruisfiltering is optioneel.

Analyse-uitvoerbestanden

Lopende runs worden weergegeven in het tabblad Active (Actief). Voltooiden runs worden weergegeven in het tabblad Completed (Voltooid). DRAGEN for IDPE Dx creëert een unieke analysemap voor elke analyse, die los staat van de map met sequencinggegevens. De analysemap omvat de volgende informatie:

- 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 locatie van de analysemap wordt gespecificeerd door de instelling External Storage for Analysis Results (Externe opslag voor analyseresultaten). Raadpleeg de Softwarehandleiding Illumina Run Manager voor NextSeq 550Dx (documentnr. 200025239) voor meer informatie over de instelling External Storage for Analysis Results (Externe opslag voor analyseresultaten).

Op het scherm Run Details biedt het veld External Location (Externe locatie) het pad voor sequencinggegevens. De unieke naam van de analysemap wordt weergegeven in het veld Analysis Output Folder (Uitvoermap van analyses) op het scherm Run Details. Welke bestanden precies worden gegenereerd, hangt af van de gebruikte analyseworkflow. De toepassing genereert de volgende uitvoerbestanden van analyses.

OPMERKING Als er een fout optreedt bij het beperken van de maximale padlengte van het bestand bij het openen van uitvoerbestanden van analyses, probeer het bestand dan naar een kortere padlocatie te verplaatsen of gebruik een andere methode om het bestand te openen.

Uitvoerbestand	Beschrijving
Overzichtsrapport variant (*.pdf)	Bevat een samenvatting van de bestandsinformatie, softwareversies, monsterinformatie, statistieken van het bepalingsniveau en SNV, invoegingen, verwijderingen en dekkingsoverzichten. Alleen de kiemlijn- en somatische workflows produceren een variantenrapport.
FASTQ (*.fastq.gz of *.fastq.ora)	Tussentijdse bestanden die basebepalingen met kwaliteitsscores bevatten. FASTQ-bestanden zijn de primaire invoer voor de uitlijningsstap. Wanneer ORA- compressie is geselecteerd, wordt de bestandsextensie *.fastq.ora gebruikt.
Uitlijning BAM-bestanden (*.bam)	Bevat uitgelijnde bepalingen voor een bepaald monster.

Uitvoerbestand	Beschrijving
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 niet-geïndexeerde 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
- De Phred-kwaliteitscores 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=AAAAAAAAA9#:<#<;<<<????#=#
```

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, bepalingvolgorde, bepalingkwaliteit, 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: Tekensreeks met overeenkomende beschrijvende elementen.
- XN: amplicon-naamtag, die de amplicon-ID registreert die bij de bepalingswaarde hoort

BAM-indexbestanden (*.bam.bai) bieden een index van het overeenkomstige 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, 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	Beschrijving
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.

Kopregel	Beschrijving
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.
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 Kiemlijnworkflow VCF-bestandsannotaties

Kopregel	Beschrijving
FILTER (Filter)	<p>Als alle filters voldoen, wordt PASS in de filterkolom geschreven. Mogelijke FILTER-invoeren zijn:</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 de 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

Kopregel	Beschrijving
FILTER (Filter)	<ul style="list-style-type: none"> • 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
INFO (Informatie)	<p>Mogelijke INFO-invoergegevens zijn:</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.

Kopregel	Beschrijving
FORMAT (Indeling)	<p>In de indelingskolom staan de velden gescheiden door dubbele punten. Bijvoorbeeld: GT:GQ. Beschikbare velden 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. • SQ: Somatische kwaliteit.
SAMPLE (Monster)	<p>In de kolom Sample (Monster) staan de waarden die in de kolom FORMAT (Indeling) zijn opgegeven.</p>

Tabel 3 Somatische workflow VCF-bestandsannotaties

Kopregel	Beschrijving
FILTER (Filter)	<p>Als alle filters voldoen, wordt PASS in de filterkolom geschreven. Mogelijke FILTER-invoeren zijn:</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 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 • 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	Beschrijving
INFO (Informatie)	<p>Mogelijke INFO-invoergegevens zijn:</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 bepalingsmappingkwaliteiten.• 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	Beschrijving
FORMAT (Indeling)	<p>In de indelingskolom staan de velden gescheiden door dubbele punten. Bijvoorbeeld: GT:GQ. Beschikbare velden 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 fasering, 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.
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.

Analyse opnieuw in de wachtrij plaatsen

U kunt de analyse opnieuw in de wachtrij plaatsen als de analyse is stopgezet, als de analyse is mislukt of als u een run opnieuw wilt analyseren met andere instellingen. Voer de volgende stappen uit om de analyse opnieuw in de wachtrij te plaatsen:

1. Selecteer in het scherm Run het tabblad Completed (Voltooid) en selecteer vervolgens de naam van de run om opnieuw te analyseren.
Als het opnieuw in de wachtrij plaatsen van de analyse eerder is uitgevoerd, selecteert u de runnaam van de bovenliggende run.
2. Selecteer in het scherm Run Details, na Sequencing Information (Sequencinginformatie), **Requeue Analysis** (Analyse opnieuw in de wachtrij plaatsen).
3. Selecteer een optie:
 - Requeue analysis with no changes (Analyse opnieuw in de wachtrij plaatsen zonder wijzigingen)
 - Edit run settings and requeue analysis (Runinstellingen en in de wachtrij plaatsen van analyse bewerken)
 - Requeue analysis with a different application (In de wachtrij plaatsen van analyse met een andere toepassing)
4. Bevestig dat de locatie waar de sequencinggegevens zich momenteel bevinden, wordt opgegeven in het veld **Sequencing data file path** (Bestandspad sequencinggegevens).

OPMERKING Het pad naar de sequencinggegevens moet overeenkomen met het pad in de instelling Externe opslag voor analyseresultaten. Raadpleeg de Softwarehandleiding Illumina Run Manager voor NextSeq 550Dx (documentnr. 200025239) voor informatie over het wijzigen van het externe opslagpad.

5. Voer een reden voor de heranalyse in.
6. Selecteer **Requeue Analysis** (Analyse opnieuw in de wachtrij plaatsen).
7. Bewerk de gewenste wijzigingen in de Runinstellingen, Monstergegevens en Analyse-instellingen.
8. Selecteer **Save** (Opslaan). De analyse wordt uitgevoerd met de huidige analyseparameters.

Technische ondersteuning

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

Website: www.illumina.com

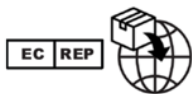
E-mail: techsupport@illumina.com

Veiligheidsinformatiebladen (SDS) – 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



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

Australische sponsor

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

VOOR GEBRUIK BIJ IN-VITRODIAGNOSTIEK.

© 2023 Illumina, Inc. Alle rechten voorbehouden.

illumina[®]