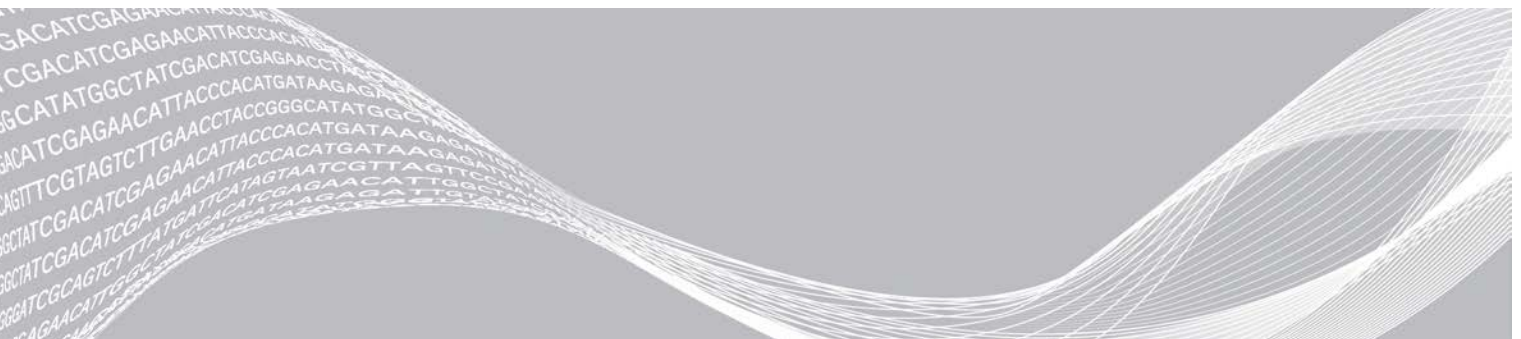


Somatic Variant-analysemodule in Local Run Manager

Workflowhandleiding voor NextSeq 550Dx

BESTEMD VOOR IN-VITRODIAGNOSTIEK

Overzicht	3
Runinformatie invoeren	4
Analysemethoden	6
Run- en monstergegevens weergeven	7
Analyserapport	8
Uitvoerbestanden van analyse	9
Basebepaling en indexdiversiteit	16
Revisiegeschiedenis	17
Technische ondersteuning	18



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 de hierin beschreven producten en voor geen enkel ander doel. Dit document en de inhoud ervan mogen niet worden gebruikt of gedistribueerd voor welk ander doel dan ook en/of op een andere manier worden gecommuniceerd, geopenbaard of gereproduceerd zonder de voorafgaande schriftelijke toestemming van Illumina. Illumina geeft door middel van dit document geen licenties onder haar patent, handelsmerk, auteursrecht of gewoonterechten noch soortgelijke rechten van derden door.

De instructies in dit document moeten strikt en uitdrukkelijk worden opgevolgd door gekwalificeerd en voldoende opgeleid personeel om een correct en veilig gebruik van de hierin beschreven producten te waarborgen. Alle inhoud van dit document moet volledig worden gelezen en begrepen voordat dergelijke producten 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 PRODUCT.

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).

© 2021 Illumina, Inc. Alle rechten voorbehouden.

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

Overzicht

De Somatic Variant-module in Local Run Manager is bedoeld voor gebruik met de Illumina TruSeq Custom Amplicon Kit Dx assay en de NextSeq 550Dx. Bij gebruik met de Somatic Variant-module is de analyse bedoeld voor het prepareren van bibliotheken die worden gebruikt voor de DNA-sequencing uit formaline-vaste, in paraffine ingebedde weefsels (FFPE). Tijdens de analyse worden somatische mutaties bij lage variantfrequenties gedetecteerd.

De analysemodule controleert korte regio's van geamplificeerd DNA, of amplicons, op varianten. Door gerichte sequencing van amplicons is een hoge dekking van bepaalde regio's in een groot aantal monsters mogelijk. De analysemodule voert secundaire analyses uit en genereert rapporten van sequencingruns met behulp van een tweestrengsaanpak waarbij gebruik wordt gemaakt van voorwaartse en achterwaartse oligo pools. Raadpleeg de bijsluiter *TruSeq Custom Amplicon Kit Dx (document # 1000000029772)*.

De Somatic Variant-analysemodule vereist verbruiksgoederen voor sequencing van 300 cycli. Raadpleeg voor meer informatie de bijsluiter *NextSeq 550Dx High Output Reagent Kit v2* of *NextSeq 550Dx High Output Reagent Kit v2.5*.

Toelichting op deze handleiding

In deze handleiding vindt u instructies voor het instellen van runparameters voor sequencing en analyse voor de Somatic Variant-analysemodule. Zie de *NextSeq 550Dx-instrumenthandleiding (documentnr. 1000000009513)* voor meer informatie over het dashboard en de systeeminstellingen van Local Run Manager.

Local Run Manager weergeven

De interface van Local Run Manager kan worden weergegeven in NextSeq 550Dx Operating Software (NOS) of met een webbrowser. De ondersteunde webbrowser is Chromium.



OPMERKING

Als u een niet-ondersteunde browser gebruikt, moet u de ondersteunde browser downloaden wanneer dat wordt gevraagd bij het bericht 'Confirm Unsupported Browser' (Niet-ondersteunde browser bevestigen). Selecteer **hier** om de ondersteunde versie van Chromium te downloaden.

Weergeven op de instrumentmonitor

- 1 Als u de interface van Local Run Manager op de instrumentmonitor wilt weergeven, selecteert u een van de volgende opties:
 - ▶ Selecteer op de NOS-startpagina de optie **Local Run Manager**.
Klik op de X in de rechterbovenhoek om terug te keren naar NOS wanneer u klaar bent.
 - ▶ Selecteer het pictogram NOS minimaliseren, open de Chromium-webbrowser die op het instrument staat en voer **http://localhost** in de adresbalk in.
Alleen admin users (beheerders) kunnen NOS minimaliseren.

Weergeven vanaf een netwerkcomputer

- 1 Open een Chromium-webbrowser op een computer met toegang tot hetzelfde netwerk als het instrument en maak verbinding met behulp van het IP-adres of de naam van het instrument.
Bijvoorbeeld: **http://myinstrument**.

Runinformatie invoeren

Parameters instellen

- 1 Log in bij Local Run Manager.
- 2 Selecteer **Create Run** (Run aanmaken) en selecteer **Somatic Variant** (Somatische variant).
- 3 Voer een runnaam in die de run van de sequencing tot en met de analyse identificeert.
Gebruik alfanumerieke tekens, spaties, onderstrepingstekens of streepjes.
- 4 **[Optioneel]** Voer een runbeschrijving in om de run te kunnen identificeren.
Gebruik alfanumerieke tekens, spaties, onderstrepingstekens of streepjes.
- 5 Selecteer het aantal monsters en de indexset in de vervolgkeuzelijst.
Overweeg de volgende informatie wanneer u een keuze maakt.
 - ▶ De vervolgkeuzelijst bevat hoeveelheden van monsters met een indexset. Bijvoorbeeld: 24-Set 1 geeft aan dat er 24 monsters moeten worden getest, met indexen uit Indexset 1.
 - ▶ De nummers van de indexsets verwijzen naar verschillende sets met i5-indexen. Set 1 en set 2 bieden allebei indexdiversiteit. Er worden twee indexsets aangeboden om uitputting van een enkele set te voorkomen.
 - ▶ Kies het aantal monsters dat het dichtst ligt bij het aantal monsters dat u gaat testen. Als het exacte aantal monsters niet in de lijst staat, selecteert u het aantal dat het dichtst bij is, maar minder dan het aantal dat u gaat testen. Als u bijvoorbeeld 18 monsters wilt testen, selecteert u 16 monsters.
 - ▶ Monsterputjes en indexcombinaties die aan de eisen voor indexdiversiteit voldoen, zijn groen gemarkeerd. Als u andere putjes en indexcombinaties selecteert, krijgt u bij het opslaan van de run een melding als niet aan de vereisten voor indexdiversiteit is voldaan.

Manifestbestanden voor de run importeren

- 1 Zorg ervoor dat de manifesten die u wilt importeren beschikbaar zijn op een toegankelijke netwerklocatie of op een USB-stick.
- 2 Selecteer **Import Manifests** (Manifesten importeren).
- 3 Blader naar het manifestbestand en selecteer de manifesten die u wilt toevoegen.



OPMERKING

Als u manifestbestanden beschikbaar wilt maken voor alle runs die de Somatic Variant-analysemodule gebruiken, voegt u manifesten toe via de module-instellingen. Voor deze functie zijn machtigingen op beheerdersniveau vereist. Raadpleeg voor meer informatie de *NextSeq 550Dx-instrumenthandleiding* (documentnr. 100000009513).


Monsters voor de run specificeren

Geef monsters op voor de run via een van de opties en de bijbehorende instructies.

- ▶ **Enter samples manually** (Monsters handmatig invoeren): gebruik de lege tabel in het scherm Create Run (Run aanmaken).
- ▶ **Import samples** (Monsters importeren): navigeer naar een extern bestand met een koma-gescheiden (*.csv) indeling. In het scherm 'Create Run' (Run aanmaken) kan een sjabloon worden gedownload.

Nadat u de monstertabel hebt gevuld, kunt u de monsterinformatie naar een extern bestand exporteren. Gebruik het bestand als referentie wanneer u bibliotheken voorbereidt of importeert het bestand voor een andere run.

Monsters handmatig invoeren

- 1 Voer in het veld Sample Name (Monsternaam) een unieke naam voor het monster in.
Gebruik alfanumerieke tekens, streepjes of onderstrepingstekens.
De naam van het monster wordt automatisch in het overeenkomstige monsterputje in de andere pool ingevoerd.
- 2 **[Optioneel]** Klik voor positieve of negatieve controlemonsters met de rechtermuisknop en selecteer het controletype.
De controle in het ene monsterputje wordt automatisch in het overeenkomstige putje in de andere pool met dezelfde controle ingevoerd.
- 3 **[Optioneel]** Voer een monsteromschrijving in het veld Sample Description (Monsterbeschrijving) in.
Gebruik alfanumerieke tekens, streepjes of onderstrepingstekens.
De omschrijving van het monster wordt automatisch in het overeenkomstige monsterputje in de andere pool ingevoerd.
Monsteromschrijvingen zijn gekoppeld aan een monster-ID. Monsteromschrijvingen worden overschreven als dezelfde monster-ID in een latere run opnieuw wordt gebruikt.
- 4 Selecteer een Index 1-adapter in de Index 1 (i7)-vervolgkeuzelijst.
Wanneer u voorgestelde monsterputjes gebruikt, vult de software automatisch i7- en i5-indexadapters in die voldoen aan de vereisten voor indexdiversiteit. Als het exacte aantal monsters dat u wilt testen niet in de lijst staat, moet u indexadapters voor extra putjes selecteren. Als u indexen moet kiezen voor extra putjes of als u niet de aanbevolen combinaties van indexadapters gebruikt, moet u, voordat u indexen kiest, eerst *Basebepaling en indexdiversiteit op pagina 16* lezen.
- 5 Selecteer een Index 2-adapter in de Index 2 (i5)-vervolgkeuzelijst.
- 6 Selecteer een manifestbestand in de vervolgkeuzelijst Manifest.
Monsters in pool A vereisen een ander manifest dan monsters in pool B.
- 7 Kies een optie om de plaatindeling te bekijken, af te drukken of op te slaan als referentie voor het prepareren van bibliotheken:
 - ▶ Selecteer het pictogram  **Print** (Afdrukken) om de indeling van de plaat weer te geven. Selecteer **Print** (Afdrukken) om de plaatindeling af te drukken.
 - ▶ Selecteer **Export** (Exporteren) om monsterinformatie naar een extern bestand te exporteren. Controleer of de gegevens van het manifest en de monsters correct zijn. Onjuiste informatie kan de resultaten beïnvloeden.
- 8 Selecteer **Save Run** (Run opslaan).

Monsters importeren

- 1 Selecteer **Import Samples** (Monsters importeren) en blader naar de locatie van het monsterinformatiebestand. U kunt twee soorten bestanden importeren.
 - ▶ Selecteer **Template** (Sjabloon) in het scherm 'Create Run' (Run aanmaken) om een nieuwe plaatindeling te maken. Het sjabloonbestand bevat de juiste kolomkoppen voor importeren. Voer in elke kolom monsterinformatie in voor de monsters in de run. Verwijder voorbeeldinformatie in ongebruikte cellen en sla het bestand op.

- ▶ Gebruik een bestand met monsterinformatie dat is geëxporteerd uit de Somatic Variant-module met de exporteerfunctie.
- 2 Selecteer het pictogram  **Print** (Afdrukken) om de indeling van de plaat weer te geven.
- 3 Selecteer **Print** (Afdrukken) om de indeling van de plaat af te drukken als referentie voor het voorbereiden van bibliotheken.
- 4 **[Optioneel]** Selecteer **Export** (Exporteren) om monsterinformatie naar een extern bestand te exporteren. Controleer of de gegevens van het manifest en de monsters correct zijn. Onjuiste informatie kan de resultaten beïnvloeden.
- 5 Selecteer **Save Run** (Run opslaan).

Een run bewerken

Raadpleeg voor instructies voor het bewerken van de informatie in uw run voordat u een sequencing start de *NextSeq 550Dx-instrumenthandleiding (documentnr. 1000000009513)*.

Analysemethoden

De Somatic Variant-analysemodule voert de volgende analysestappen uit en schrijft de uitvoerbestanden van de analyse naar de map Alignment.

- ▶ Voert demultiplexen van indexbepalingen uit
- ▶ Genereert FASTQ-bestanden
- ▶ Lijnt uit op een referentie
- ▶ Identificeert varianten

Demultiplexen

Met demultiplexen wordt elke indexbepalingssequentie met de voor de run opgegeven indexsequenties vergeleken. In deze stap worden geen kwaliteitswaarden overgenomen.

Indexbepalingen worden geïdentificeerd aan de hand van de volgende stappen:

- ▶ Monsters worden genummerd vanaf 1 in 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 toegewezen aan een monster wanneer de indexsequentie exact overeenkomt of wanneer er maximaal één discrepantie per indexbepaling is.

FASTQ-bestand genereren

Na het demultiplexen genereert de software tussenliggende analysebestanden in de FASTQ-bestandsindeling, wat een tekstindeling is die wordt gebruikt om sequenties weer te geven. FASTQ-bestanden bevatten bepalingen voor elk monster en de bijbehorende kwaliteitsscores. Clusters die niet door het filter zijn gekomen, worden uitgesloten.

Elk FASTQ-bestand bevat bepalingen van slechts één monster. De naam van dat monster is opgenomen in de FASTQ-bestandsnaam. FASTQ-bestanden zijn de primaire invoer voor uitlijning. Per monster en per oligo pool worden acht FASTQ-bestanden gegenereerd, vier van Bepaling 1 en vier van Bepaling 2, wat resulteert in een totaal van 16 FASTQ-bestanden per monster.

Uitlijning

Tijdens de uitlijningsstap lijnt het gebandeerde Smith-Waterman-algoritme clusters van elk monster uit met amplicon-sequenties die in het manifestbestand zijn opgegeven.

Het gebandeerde Smith-Waterman-algoritme voert semiglobale sequentie-uitlijningen uit om gelijksoortige gebieden tussen twee sequenties te bepalen. In plaats van de totale sequentie te vergelijken, vergelijkt het Smith-Waterman-algoritme segmenten van alle mogelijke lengtes.

Elke paired-end-bepaling wordt geëvalueerd wat betreft uitlijning met de relevante probesequenties voor die bepaling.

- ▶ Bepaling 1 wordt geëvalueerd aan de hand van de omgekeerde aanvulling van de Downstream Locus-Specific Oligos (DLSO).
- ▶ Bepaling 2 wordt geëvalueerd aan de hand van de Upstream Locus-Specific Oligos (ULSO).
- ▶ Indien het begin van een afgelezen waarde overeenkomt met een probesequentie en niet meer dan drie verschillen bevat (discrepancies of verschuivingen door leidende indels), wordt de volledige lengte van de afgelezen waarde uitgelijnd met het amplicondoel voor die sequentie.
- ▶ Gegeven de chemie van de assay worden indels binnen de DLSO en ULSO niet waargenomen.

Uitlijningen worden uit uitlijningsresultaten gefilterd op basis van discrepantiepercentages over ofwel het gebied van belang ofwel het volledige amplicon, afhankelijk van de lengte van het amplicon. Gefilterde uitlijningen worden in uitlijningsbestanden weggeschreven als niet-uitgelijnd en worden niet gebruikt bij het aanroepen van varianten.

Variantbepaling

De Pisces-variantbepaler identificeert varianten die met een lage frequentie in het DNA-monster voorkomen.


De Pisces-variantbepaler identificeert SNV's, MNV's en kleine indels in drie stappen:

- ▶ Overweegt elke positie in het referentiegenoom afzonderlijk
- ▶ Telt basen op de gegeven positie voor uitgelijnde bepalingen die de positie overlappen
- ▶ Berekent een variantscore die de kwaliteit van de bepaling meet met behulp van het Poisson-model. Varianten met een kwaliteitsscore lager dan Q30 worden uitgesloten.

Varianten worden eerst voor elke pool apart bepaald. Vervolgens worden de varianten uit elke pool vergeleken en gecombineerd tot één uitvoerbestand. Als een variant in beide pools voorkomt en aan alle filters voldoet die zijn vermeld in *VCF-bestandsannotaties op pagina 12*, wordt de variant in het variantbepalingsbestand (VCF) als PASS (Geslaagd) gemarkeerd.

Run- en monstergegevens weergeven

- 1 Klik in het dashboard van Local Run Manager op de runnaam.
- 2 Bekijk op het tabblad Run Overview (Runoverzicht) de metrische gegevens van de sequencing-run.
- 3 **[Optioneel]** Klik op het pictogram **Copy to Clipboard**  (Naar klembord kopiëren) om het pad naar de uitvoermap voor de run te kopiëren.
- 4 Klik op het tabblad Sequencing Information (Sequencing-informatie) om de runparameters en de informatie over de verbruiksartikelen te bekijken.

- 5 Klik op het tabblad Samples and Results (Monsters en resultaten) om de locatie van het analyserapport weer te geven.
 - ▶ Als de analyse is herhaald, opent u het vervolgkeuzemenu Select Analysis (Analyse selecteren) en selecteert u de juiste analyse.
- 6 Klik op het pictogram **Copy to Clipboard**  (Naar klembord kopiëren) om het pad naar de analysemap te kopiëren.

Raadpleeg voor meer informatie over de tabbladen Run Overview (Runoverzicht) en Sequencing Information (Sequencinginformatie), en hoe u een analyse opnieuw kunt uitvoeren, de *NextSeq 550Dx-instrumenthandleiding* (documentnr. 1000000009513).

Analyserapport

De analyseresultaten staan in een samenvatting op het tabblad Samples and Results (Monsters en resultaten) en in een samenvattend rapport in de map Alignment (Uitlijning). Voor elk monster is ook een rapport in PDF-indeling beschikbaar.

Informatie op het tabblad Samples and Results (Monsters en resultaten)

- 1 Klik op een monster in de lijst om het monsterrapport te openen.

Tabel 1 Informatie over run en monster

Kolomkop	Omschrijving
Run Status (Runstatus)	Geeft aan of de sequencing-run geslaagd of mislukt is.
Total Yield (GB) (Totaal resultaat (GB))	Aantal basen die in de sequencing-run zijn bepaald. Toont de drempelwaarde en de status geslaagd of mislukt.
% ≥ Q30	Het percentage bepalingen in de sequencing-run met een kwaliteitsscore van 30 (Q30) of hoger. Toont de drempelwaarde en de status geslaagd of mislukt.
Sample Name (Monsternaam)	De monsternaam die bij het aanmaken van de run is opgegeven.
Total PF Reads (Totaal PF-bepalingen)	Het totale aantal bepalingen dat door het filter komt.
Read 1% ≥ Q30 (Bepaling 1% ≥ Q30)	Het percentage bepalingen in Bepaling 1 met een kwaliteitsscore van 30 (Q30) of hoger voor het monster.
Read 2% ≥ Q30 (Bepaling 2% ≥ Q30)	Het percentage bepalingen in Bepaling 2 met een kwaliteitsscore van 30 (Q30) of hoger voor het monster.
Autosome Call Rate (Bepalingspercentage van autosomen)	Het aantal genomische posities in de autosomen (chromosomen 1 tot en met 22) dat aan een vooraf bepaalde drempelwaarde voor de betrouwbaarheidsgraad voldoet, gedeeld door het totale aantal onderzochte autosomale genomische posities. Het bepalingpercentage wordt per monster beschreven en gerapporteerd als 1 min (het aantal autosomale posities met onvolledige bepalingen gedeeld door het totale aantal autosomale posities waarvan de sequentie is bepaald).

Tabel 2 Informatie in monsterrapport

Kolomkop	Omschrijving
Sample (Monster)	De monsternaam die bij het aanmaken van de run is opgegeven.
Report Date (Rapportdatum)	Datum waarop het rapport werd gegenereerd.
Sample Information (Monsterinformatie)	De monster-ID die werd opgegeven bij het aanmaken van de run, het totaal aantal bepalingen die door de filter in het monster zijn gekomen, het bepalingpercentage voor het monster met een kwaliteitsscore van 30 (Q30) of hoger en het autosomale bepalingpercentage.
Amplicon Summary (Ampliconsamenvatting)	Totaal aantal gesequencete amplicongebieden en de totale lengte in basenparen van gesequencete amplicons in de doelgebieden, voor het monster in pool A en pool B en het voor elke pool gebruikte manifestbestand. Het manifestbestand specificiert het referentiegenoom en de doelreferentiegebieden die in de uitlijningsstap worden gebruikt.
Read Level Statistics (Statistieken van bepalingsniveaus)	Aantal en percentage van bepalingen voor het monster die elke positie in de referentie dekken voor Bepaling 1 en Bepaling 2 in Pool A en Pool B.
Variants Summary (Overzicht van varianten)	Aantal gedetecteerde SNV's, invoegingen en verwijderingen voor het monster dat voldoet aan de voorgestelde waarden om te bepalen of de kwaliteitsresultaten binnen een aanvaardbaar bereik liggen,
Coverage Summary (Dekkingsamenvatting)	Het totale aantal uitgelijnde basen gedeeld door de grootte van het doelgebied en het percentage amplicongebieden met dekkingswaarden die groter zijn dan de lage dekkingsdrempel van $0,2 * \text{amplicon gemiddelde dekking}$ voor het monster in Pool A en Pool B.
Coverage Plots (Dekkingsdiagrammen)	De grafiek Coverage by Amplicon Region (Dekking per amplicongebied) toont de dekking binnen de amplicongebieden voor het monster. Gebieden met dekkingswaarden die lager zijn dan de dekkingsdrempel, zijn rood gemarkeerd. Het gemiddelde van alle waarden wordt aangegeven door een oranje lijn. Er is een diagram beschikbaar voor de dekking van Pool A en Pool B.
Software Versions (Softwareversies)	Softwareversies toen het monster werd gesequencet. Omvat de NextSeq 550Dx-besturingssysteemsoftware (NOS), de Local Run Manager-software, de RTA-software en de Somatic Variant-moduleversie.

Uitvoerbestanden van analyse

De volgende uitvoerbestanden van de analyse worden gegenereerd voor de Somatic Variant-analysemodule en bevatten analyseresultaten voor uitlijning en bepaling van varianten. De uitvoerbestanden van de analyse bevinden zich in de map Alignment.

Bestandsnaam	Omschrijving
Demultiplexen (*.txt)	Tussenliggende bestanden met samenvattende demultiplexresultaten.
FASTQ (*.fastq.gz)	Tussenliggende bestanden die basebepalingen met kwaliteitsscores bevatten. FASTQ-bestanden zijn de primaire invoer voor de uitlijningsstap.
Uitlijningsbestanden in de BAM-bestandsindeling (*.bam)	Bevat uitgelijnde bepalingen voor een bepaald monster.
Per-pool-variantbepalingsbestanden in de VCF-bestandsindeling (*.vcf)	Bevat varianten die op elke positie worden bepaald vanuit de voorwaartse of de achterwaartse pool.
Variantbepalingsbestanden met genome VCF-indeling (*.genome.vcf.gz)	Bevat het genotype voor elke positie, of het nu als variant of als referentie wordt aangeropen.

Bestandsnaam	Omschrijving
Consensusvariantbepalingsbestanden met VCF-bestandsindeling (*.vcf.gz)	Bevat varianten die op elke positie vanuit beide pools worden bepaald.
AmpliconCoverage_M1.tsv	Bevat informatie over de dekking per amplicon en per monster voor elk verstrekt manifest. M# staat voor het manifestnummer.

Bestandsindeling voor demultiplexen

Bij demultiplexen wordt de aan elk cluster gekoppelde indexsequentie gelezen om te bepalen van welk monster het cluster afkomstig is. De toewijzing van clusters aan monsternummers wordt weggeschreven naar een demultiplexbestand (*.demux) voor elke tegel van de stroomcel.

De naamgeving van het demultiplexbestand is **s_1_X.demux**, waarbij X het nummer van de tegel is.

Demultiplexbestanden beginnen met een kopregel:

- ▶ Versie (geheel getal van 4 bytes), momenteel 1
- ▶ Clustertelling (geheel getal van 4 bytes)

De rest van het bestand bestaat uit monsternummers voor elk cluster uit de tegel.

Wanneer de demultiplexstap is voltooid, genereert de software een demultiplexbestand met de naam **DemultiplexSummaryF1L1.txt**.

- ▶ In de bestandsnaam staat **F1** voor het stroomcelnummer.
- ▶ In de bestandnaam staat **L1** voor het baannummer.
- ▶ Demultiplexen resulteert in een tabel met 1 rij per tegel en 1 kolom per monster, inclusief monster 0.
- ▶ De meest voorkomende sequenties in indexbepalingen.

FASTQ-bestandsindeling

FASTQ is een tekstbestandsindeling die basepalings en kwaliteitswaarden per bepaling bevat. Elke record bevat 4 regels:

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

De identificatie is als volgt ingedeeld:

@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=AAAAAAAAAA9#:<#<;<<<?????#=#
```

BAM-bestandsindeling

Een BAM-bestand (*.bam) is de gecomprimeerde binaire versie van een SAM-bestand dat wordt gebruikt om uitgelijnde sequenties tot 128 Mb weer te geven. De bestandsindelingen SAM en BAM worden uitgebreid beschreven op samtools.github.io/hts-specs/SAMv1.pdf.

BAM-bestanden hebben de naam `SampleName_S#.bam`, waarbij # het monsternummer is dat wordt bepaald door de plaats van het monster in de runlijst.

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 gekoppelde einde-uitlijning
- ▶ **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 bepalingswaarde hoort

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

VCF-bestandsindeling

Variant Call Format (VCF) is een veelgebruikte bestandsindeling die is ontwikkeld door de wetenschappelijke genomicsgemeenschap. VCF-bestanden bevatten informatie over varianten die op specifieke posities in een referentiegenoom worden aangetroffen. VCF-bestanden eindigen op het achtervoegsel `.vcf`

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.

VCF-bestandskopregels

Kopregel	Omschrijving
CHROM	Het chromosoom van het referentiegenoom. Chromosomen verschijnen in dezelfde volgorde als het FASTQ-referentiebestand.
POS	De enkele basepositie van de variant in het referentiechromosoom. Voor SNP's is deze positie de referentiebasis met de variant. Voor indels of verwijderingen is deze positie de referentiebasis onmiddellijk vóór de variant.
Id	Het rs-nummer voor de variant verkregen uit dbSNP.txt, indien van toepassing. Als er meerdere rs-nummers op deze locatie staan, 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	Omschrijving
ALT	De allelen die verschillen van de referentiebeplating. 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 score op de 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.

VCF-bestandsannotaties

Kopregel	Omschrijving
FILTER (Filter)	Als aan alle filters is voldaan, wordt PASS in de filterkolom geschreven. <ul style="list-style-type: none"> • LowDP: Toegepast op locaties met een dekkingsdiepte van minder dan 450 x in beide pools. Voor ampliconposities die zowel door de voorwaartse als door de achterwaartse bepaling worden gedekt, komt dit overeen met een dekking van 900 x enkele bepaling. • LowGQ: De genotyperingskwaliteit (GQ) ligt onder een grenswaarde. • q30: Kwaliteitsscore < 30. • LowVariantFreq: De variantfrequentie is lager dan de opgegeven drempel. • PB: Systematische fout van probepool. Variant niet gevonden of met lage frequentie gevonden in één of twee probepools. • R3x6: Het aantal aangrenzende herhalingen (van lengte 1 tot 3 bp) op de variantbepalingen \geq 6. • SB: De systematische fout van de streng is groter dan de bepaalde drempel.
INFO (Informatie)	Mogelijke vermeldingen in de kolom INFO zijn: <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 genoemde genotypen. • CD: Een vlag die aangeeft dat de SNP voorkomt in het coderende gebied van ten minste 1 RefGene-vermelding. • DP: De diepte (aantal basepalingen uitgelijnd op een positie en gebruikt bij het bepalen van varianten). • Exon: Een door komma's gescheiden lijst van exongebieden die uit RefGene zijn gelezen. • FC: Functioneel gevolg. • GI: Een door komma's gescheiden lijst van gen-ID's die uit RefGene zijn gelezen. • QD: Variantvertrouwen/kwaliteit per diepte. • TI: Een door komma's gescheiden lijst van transcript-ID's die uit RefGene zijn gelezen.

Kopregel	Omschrijving
FORMAT (Indeling)	<p>In de indelingskolom staan de velden gescheiden door dubbele punten. Bijvoorbeeld: GT:GQ. De lijst van verstrekte velden hangt af van de gebruikte variantbepaler. Beschikbare velden zijn:</p> <ul style="list-style-type: none"> • AD: Invoer van de vorm X, Y, waarbij X het aantal referentie-bepalingen is en Y het aantal alternatieve bepalingen. • DP: Geschatte bepalingstiepte; bepalingen met MQ=255 of met slechte paren worden gefilterd. • GQ: Kwaliteit van het genotype. • GQX: Kwaliteit van het genotype. GQX is het minimum van de GQ-waarde en de kolom QUAL. In het algemeen zijn deze waarden vergelijkbaar. Door het minimum te nemen wordt GQX de conservatievere maatstaf voor genotypekwaliteit. • 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. • NL: Ruisniveau; een schatting van de basebepalingsruis op deze positie. • PB: Systematische fout van probepool. Waarden dicht bij 0 duiden op meer systematische fouten van één probepool en minder vertrouwen in een variantbepaling. • SB: Systematische fout van streng op deze positie. Grotere negatieve waarden wijzen op minder systematische fouten. Waarden dicht bij 0 wijzen op meer systematische fouten. • VF: Variantfrequentie; het bepalingpercentage dat het alternatieve allel ondersteunt.
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) zijn VCF v4.1-bestanden die een reeks conventies volgen om alle locaties binnen het genoom in een redelijk compact formaat weer te geven. De gVCF-bestanden (*.genome.vcf.gz) bevatten alle locaties binnen de betrokken regio 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.

Zie voor meer informatie sites.google.com/site/gvcftools/home/about-gvcf.

Per-Pool- en Consensus VCF-bestanden

De Somatic Variant-workflow genereert 2 sets variantbepalingsbestanden.

- ▶ **Per-pool VCF-bestanden**: bevat varianten die in de voorwaartse of achterwaartse pool worden bepaald. Per-pool-bestanden worden in de map VariantCallingLogs opgeslagen.
- ▶ **Consensus VCF-bestanden**: bevatten varianten die vanuit beide pools worden bepaald. Consensus-bestanden worden in de map Alignment (Uitlijning) opgeslagen.

Per-Pool- en Consensus-bestanden omvatten zowel VCF- (*.vcf) als gVCF-bestanden (*.genome.vcf) en gebruiken de volgende naamgevingsconventie, waarbij S# staat voor de volgorde waarin het monster in de run is opgenomen:

- ▶ **Rapporten voor alle locaties**: SampleName_S#.genome.vcf
- ▶ **Rapporten voor alleen varianten**: SampleName_S#.vcf

De software vergelijkt de Per-Pool VCF-bestanden en combineert de gegevens op elke positie om een Consensus VCF-bestand voor het monster aan te maken.

Variantbepalingen uit elke pool worden samengevoegd tot Consensus VCF-bestanden volgens de volgende criteria.

Criteria	Resultaat
Een referentie bepaling in elke pool	Referentie bepaling
Een referentie bepaling in 1 pool en een variant bepaling in de andere pool	Gefilterde variant bepaling
Overeenkomende variant bepalingen met vergelijkbare frequenties in elke pool	Variant bepaling
Overeenkomende variant bepalingen met sterk afwijkende frequenties in elke pool	Gefilterde variant bepaling
Niet-overeenkomende variant bepalingen in elke pool	Gefilterde variant bepaling

Metrische gegevens uit elke pool worden samengevoegd met behulp van de volgende waarden.

Metrisch gegeven	Waarde
Diepte	Toevoeging van diepten uit beide pools
Variantfrequentie	Totaal aantal varianten gedeeld door totale dekkingsdiepte
Kwaliteitsscore	Minimumwaarde van beide pools

Bestand met amplicon-dekking

Voor elk manifestbestand wordt een bestand met amplicon-dekking gegenereerd. Het M-nummer in de bestandsnaam staat voor het manifestnummer.

Elk bestand bevat een kopregelrij met de monster-ID's die bij het manifest horen. Het bestand bevat de volgende informatie.

- ▶ De doel-ID zoals die in het manifest staat vermeld.
- ▶ De dekkingsdiepte van bepalingen die door het filter komen.

Aanvullende uitvoerbestanden

De volgende uitvoerbestanden bevatten aanvullende informatie of een samenvatting van de runresultaten en analysefouten. Hoewel deze bestanden niet nodig zijn voor de beoordeling van de analyseresultaten, kunnen ze worden gebruikt voor het oplossen van problemen. Alle bestanden bevinden zich in de map Alignment, tenzij anders aangegeven.

Bestandsnaam	Omschrijving
AnalysisLog.txt	Verwerkingslogboek waarin elke stap wordt beschreven die is uitgevoerd tijdens de analyse van de huidige runmap. Dit bestand bevat geen foutmeldingen. Bevindt zich in de map Alignment.
AnalysisError.txt	Verwerkingslogboek waarin een lijst is opgenomen van eventuele fouten die tijdens de analyse zijn opgetreden. Dit bestand is leeg als er geen fouten zijn opgetreden. Bevindt zich in de map Alignment.
DemultiplexSummaryF1L1#.txt	Rapporteert demultiplexresultaten in een tabel met 1 rij per tegel en 1 kolom per monster. Het # staat voor baan 1, 2, 3 of 4 van de stroomcel. Bevindt zich in de map Alignment.
AmpliconRunStatistics.xml	Bevat samenvattende statistieken die specifiek zijn voor de run. Bevindt zich in de map Alignment.

Analysemap

De analysemap bevat de bestanden die worden gegenereerd door de Local Run Manager-software.


De relatie tussen de uitvoer- en de analysemap kan als volgt worden samengevat:

- ▶ Tijdens de sequencing vult Real-Time Analysis (RTA) de uitvoermap met bestanden die tijdens de beeldanalyse, basebepaling en kwaliteitsscore zijn gegenereerd.
- ▶ RTA kopieert bestanden in realtime naar de analysemap. Nadat RTA een kwaliteitsscore heeft toegekend aan elke basis voor elke cyclus, schrijft de software het bestand RTAComplete.txt naar beide mappen.
- ▶ Wanneer het bestand RTAComplete.txt aanwezig is, begint de analyse.
- ▶ Terwijl de analyse wordt voortgezet, schrijft Local Run Manager uitvoerbestanden naar de analysemap en kopieert de bestanden terug naar de uitvoermap.

Uitlijningsmappen

Telkens wanneer die analyse wordt opgevraagd, maakt Local Run Manager een uitlijningsmap aan met de naam **Alignment_N**, waarbij N een volgnummer is.

Mapstructuur

 **Alignment:** bevat *.bam-, *.vcf-, FASTQ-bestanden en bestanden die specifiek zijn voor de analysemodule.

 **Date and Time Stamp:** datum- en tijdstempel van analyse als JJJJMMDD_UUMMSS

 AnalysisError.txt

 AnalysisLog.txt

 aggregate.report.html

 aggregate.report.pdf

 aggregate.summary.csv

 AmpliconCoverage_M#.tsv

 AmpliconRunStatistics.xml

 Sample1.genome.vcf.gz

 Sample1.coverage.csv

 Sample1.report.pdf

 Sample1.summary.csv

 Sample1.vcf.gz

 Sample1.bam

 **FASTQ**

 **Sample1**

 Sample1_L001_R1_001_fastq.gz

 **Stats**

 DemuxSummaryF1L1.txt

FastqSummaryF1L1.txt

Data

Intensities

BaseCalls

L001: bevat *.bcl-bestanden.

L001: bevat *.locs-bestanden.

RTA Logs: bevat logboekbestanden van de RTA-softwareanalyse.

InterOp: bevat binaire bestanden die worden gebruikt om metrische gegevens van de sequencing-run te rapporteren.

Logs: bevat logboekbestanden waarin de stappen worden beschreven die zijn uitgevoerd tijdens de sequencing.

RTAComplete.txt

RunInfo.xml

RunParameters.xml

Basebepaling en indexdiversiteit

Wanneer monsters worden gesequencet op het NextSeq 550Dx-instrument, wordt door basebepaling een base bepaald (A, C, G of T) voor elk cluster van een bepaalde tegel of beeldvormingsgebied op de stroomcel, bij een specifieke cyclus. Het NextSeq 550Dx-instrument maakt gebruik van sequencing met twee kanalen, waardoor slechts twee beelden nodig zijn om de gegevens voor vier DNA-basen te coderen, één van het rode kanaal en één van het groene kanaal.

De procedure voor basebepaling-indexbepalingen is anders dan voor basebepaling tijdens andere bepalingen.

Indexbepalingen moeten beginnen met ten minste één base anders dan G in een van de eerste twee cycli. Als een indexbepaling begint met twee basebepalingen van G, wordt er geen signaalintensiteit gegenereerd. Er moet een signaal aanwezig zijn in één van de eerste twee cycli om demultiplexprestatie te garanderen.

Bij het selecteren van indexen tijdens het aanmaken van runs wordt een waarschuwing voor lage diversiteit weergegeven als de indexen niet aan de diversiteitseisen voldoen. U kunt deze waarschuwing voorkomen door indexsequenties te kiezen die bij elke cyclus een signaal in beide kanalen leveren.

- ▶ Rood kanaal: A of C
- ▶ Groen kanaal: A of T

Dit basebepalingsproces garandeert nauwkeurigheid bij het analyseren van low-plexmonsters. Zie voor meer informatie over de sequenties van uw indexen de bijsluiters *TruSeq Custom Amplicon Kit Dx (documentnr. 1000000029772)*.

Tijdens het aanmaken van de run in Local Run Manager kiest u het aantal monsters dat u wilt testen. Voorgestelde indexcombinaties die voldoen aan de eisen voor indexdiversiteit worden automatisch door de software ingevuld. Hoewel u niet verplicht bent de voorgestelde indexcombinaties te gebruiken, wordt het wel aanbevolen.

Revisiegeschiedenis

Document	Datum	Omschrijving van wijziging
Documentnr. 1000000030330 v04	Augustus 2021	Adres van de gemachtigde vertegenwoordiger van de EU bijgewerkt.
Documentnr. 1000000030330 v03	April 2020	Adres van de gemachtigde vertegenwoordiger van de EU bijgewerkt. Adres van de Australische sponsor bijgewerkt.
Documentnr. 1000000030330 v02	Januari 2019	Informatie over v2.5 reagenskits toegevoegd.
Documentnr. 1000000030330 v01	Augustus 2018	Voorgeschreven markeringen bijgewerkt.
Documentnr. 1000000030330 v00	November 2017	Eerste uitgave.

Technische ondersteuning

Voor technische ondersteuning neemt u contact op met de afdeling technische ondersteuning van Illumina.

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

Telefoonnummers van klantenondersteuning van Illumina

Regio	Gratis telefoonnummer	Regionaal telefoonnummer
Noord-Amerika	+1 800 809 4566	
Australië	+1.800.775.688	
België	+32 80077160	+32 34002973
China	400.066.5835	
Denemarken	+45 80820183	+45 89871156
Duitsland	+49 8001014940	+49 8938035677
Finland	+358 800918363	+358 974790110
Frankrijk	+33 805102193	+33 170770446
Hongkong	800960230	
Ierland	+353 1800936608	+353 016950506
Italië	+39 800985513	+39 236003759
Japan	0800 111 5011	
Nederland	+31 8000222493	+31 207132960
Nieuw-Zeeland	0800 451 650	
Noorwegen	+47 800 16836	+47 21939693
Oostenrijk	+43 800006249	+43 19286540
Singapore	+1.800.579.2745	
Spanje	+34 911899417	+34 800300143
Taiwan	00806651752	
Verenigd Koninkrijk	+44 8000126019	+44 2073057197
Zweden	+46 850619671	+46 200883979
Zwitserland	+41 565800000	+41 800200442
Overige landen	+44 1799 534 000	

Veiligheidsinformatiebladen (SDS, safety data sheets) — zijn verkrijgbaar op de website van Illumina via support.illumina.com/sds.html.

Productdocumentatie — beschikbaar voor downloaden in pdf-vorm via de website van Illumina. Ga naar support.illumina.com, selecteer een product en klik vervolgens op **Documentation & Literature** (Documentatie en literatuur).



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

© 2021 Illumina, Inc. Alle rechten voorbehouden.

illumina®