

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

Brugervejledning til program

Dette dokument og dets indhold er ophavsretligt beskyttet af Illumina, Inc. og dets datterselskaber ("Illumina") og er udelukkende beregnet til kundens kontraktmæssige brug i forbindelse med anvendelsen af det produkt eller de produkter, som er beskrevet heri og til intet andet formål. Dette dokument og dets indhold må ikke bruges eller distribueres til noget andet formål og/eller på anden måde kommunikeret, offentliggøres eller reproduceres på nogen som helst måde uden forudgående, skriftligt samtykke fra Illumina. Med dette dokument udsteder Illumina ingen licens under sit patent, varemærke, sin copyright eller sædvaneret eller lignende rettigheder for nogen tredjeparter.

Anvisningerne i dette dokument skal følges nøje og fuldstændigt af kvalificerede og behørigt uddannede medarbejdere for at sikre, at det produkt eller de produkter, der er beskrevet heri, anvendes korrekt og sikkert. Alt indhold i dette dokument skal læses grundigt og forstås inden brug af produktet/produkterne.

HVIS ALLE ANVISNINGERNE HERI IKKE GENNEMLÆSES FULDT UD OG FØLGES NØJE, KAN DET MEDFØRE SKADE PÅ PRODUKTET ELLER PRODUKTERNE, SKADE PÅ PERSONER, HERUNDER BRUGERE ELLER ANDRE, OG SKADE PÅ ANDEN EJENDOM OG VIL GØRE ENHVER GARANTI GÆLDENDE FOR PRODUKTET ELLER PRODUKTERNE UGYLDIG.

ILLUMINA PÅTAGER SIG INTET ANSVAR SOM FØLGE AF FORKERT BRUG AF DET PRODUKT ELLER DE PRODUKTER, DER ER BESKREVET HERI (HERUNDER DELE HERAF ELLER SOFTWARE).

© 2023 Illumina, Inc. Alle rettigheder forbeholdes.

Alle varemærker tilhører Illumina, Inc. eller de respektive ejere. Specifikke varemærkeoplysninger er tilgængelige på www.illumina.com/company/legal.html.

Revisionshistorik

Dokument	Dato	Beskrivelse af ændring
200025238 v00	Februar 2023	Oprindelig udgivelse.

Indholdsfortegnelse

Revisionshistorik	iii
Oversigt	1
Analysemetoder	1
Opret en planlagt kørsel	5
Indstillinger	7
Manifestfil	8
Støjfiltrering (Valgfri)	9
Analyseoutput	9
FASTQ-filer	10
BAM-filer	11
VCF-filer	11
Genindsættelse i analysekø	18
Teknisk hjælp	20

Oversigt

DRAGEN for Illumina DNA Prep with Enrichment Dx-programmet (DRAGEN for IDPE Dx) bruges til at planlægge og udføre sekundær analyse af IDPE Dx-biblioteker genereret til sekventering på NextSeq 550Dx.

DRAGEN for IDPE Dx understøtter sekventering til analyse, når det bruges sammen med Illumina DNA Prep with Enrichment Dx Biblioteksklargøring, NextSeq 550Dx og Illumina DRAGEN Server for NextSeq 550Dx.

Analysemetoder

DRAGEN for IDPE Dx udfører demultipleksering, FASTQ-generering, læsning af kortlægning, tilpasning til et referencegenom og en mindre variantbestemmelse afhængigt af den valgte arbejdsgang:

- FASTQ-generering
- Kimcelle FASTQ- og VCF-generering
- Somatisk FASTQ- og VCF-generering

BEMÆRK ORA-komprimering kan bruges med alle tre arbejdsgange. DRAGEN ORA-komprimering er en helt tabsfri komprimeringssoftware, der opretter en fil med filendelsen Original Read Archive (*.ora). Ora-formatet er et referencebaseret komprimeringsformat til FASTQ-filer, der er designet til meget hurtig komprimering/dekomprimering og et højt komprimeringsforhold.

Generering af FASTQ

De sammensatte sekvenser skrives til FASTQ-filer pr. prøve. FASTQ-filerne er tekstfiler, der kun indeholder sekventeringsdata og kvalitetsscorer for én prøve. Der genereres separate FASTQ-filer pr. flowcellebane, pr. sekventeringslæsning for hver prøve. Prøvens navn, som angives under kørsels konfiguration, er inkluderet i FASTQ-filens navn. FASTQ-filer er det primære tilpasningsinput. Det første trin i FASTQ-generering er demultipleksering. Demultipleksering tildeler clustre, der passerer filteret, til en prøve, ved at sammenligne hvert indeksslæsningssekvens med de indeksssekvenser, der er angivet for kørslen. Ingen kvalitetsværdier tages i betragtning på dette trin. Indekslæsninger identificeres ved hjælp af følgende trin:

- Prøver nummereres fra og med 1 baseret på den rækkefølge, de er anført i for kørslen.
- Prøve nummer 0 er forbeholdt clustre, som ikke er tildelt en prøve.
- Clustre tildeles en prøve, når indeksssekvensen matcher præcist, eller når der er højst én uoverensstemmelse pr. indeksslæsning.

Softwaren inkluderer ORA-komprimering til at komprimere FASTQ-filer. Dette format kan aktiveres valgfrit. Ved brug af ORA-formatet (*.ora) bevares md5-kontrolsummen for FASTQ-indholdet efter en komprimerings- og dekomprimeringscyklus for at sikre tabsfri komprimering.

Kortlægning og tilpasning af DNA

Efter FASTQ-generering bliver læsningerne kortlagt og justeret til et referencegenom. Det første stadie i kortlægningen er at generere kimen fra læsningen og derefter prøve at finde eksakte match i referencegenomet. Disse resultater forbedres så ved at køre fuld Smith-Waterman-tilpasning på placeringerne med den højeste densitet af kimmatch. Denne veldokumenterede algoritme fungerer ved at sammenligne hver placering af læsningen i forhold til alle referencens kandidatplaceringer. Disse sammenligninger svarer til en matrix af potentielle tilpasninger mellem læsning og reference. Smith-Waterman genererer scorer for hver af disse kandidattilpasningsplaceringer, der bruges til at evaluere om den bedste tilpasning, der passerer gennem denne matrixcelle, når den via en nukleotidmatch eller mismatch (diagonal bevægelse), en sletning (vandret bevægelse) eller en indsættelse (lodret bevægelse). Et match mellem læsning og reference giver en bonus på scoren, og en mismatch eller indel trækker fra. Den samlede, højeste scoringslinje gennem matrixen er den tilpasning, der vælges. Algoritmen er hardwareaccelereret på DRAGENs feltprogrammerbare porttabel (FPGA-kort). Referencegenomet, der bruges i appen, oprettes fra UCSC hg19 FASTA med DRAGEN-formatet for at oprette en liftover-baseret alt-aware-hashtabel.

DRAGEN bestemmelse af kimcellevarianter

DRAGEN Germline Small Variant Caller tager tilknyttede og tilpassede DNA-læsninger som input og bestemmer enkeltnukleotidpolymorfi (SNP'er) og indsættelser eller sletninger (indeler) gennem en kombination af kolonnevis påvisning og lokal *de novo*-samling af haplotyper. For at aktivere DRAGEN Germline Small Variant Caller skal du vælge arbejdsgangen for kimcellevarianter.

Bestemmelse af kimcellevarianter bruges typisk til kimcelleprøver, hvor ploidi er kendt for at være to. De referenceområder der kan bestemmes, identificeres først med tilstrækkelig tilpasningsdækning. Inden for disse referenceområder vil en hurtig scanning af de sorterede læsninger identificere aktive områder, som er centreret om samlingen af kolonner med evidens for en variant. De aktive områder er polstret med tilstrækkelig kontekst til at dække signifikant, ikke-reference-indhold i nærheden. Hvis der er evidens for indeler, vil de aktive områder få ekstra polstring.

Tilpassede læsninger klippes inden for hver aktivt område og samles i en De Bruijn-graf. Kanterne på de klippede læsninger vægtes af observationstal, med referencesekvensen som et skelet. Efter lidt oprydning og forenkling af grafen, ekstraheres alle source-to-sink-linjer som kandidat-haplotyper. Hver haplotype er Smith-Waterman tilpasset til referencegenomet for at identificere de varianter, den repræsenterer. Dette sæt hændelser kan suppleres med en placeringsbaseret påvisning. For hvert læsning-haplotype-par estimeres sandsynligheden $P(r|H)$ for at observere læsningen, når det antages at haplotypen er den sande startprøve, ved brug af en par-skjult Markov-model (HMM).

Ved scanning efter referenceplacering over det aktive område dannes der kandidatgenotyper fra diploidkombinationer af varianthændelser (SNP'er eller indeler). For hver hændelse (herunder reference) estimeres den betingede sandsynlighed $P(r|e)$ for at observere hver overlappende læsning, som den maksimale $P(r|H)$ for haplotyper, der understøtter denne hændelse. Disse kombineres til den betingede sandsynlighed $P(r|e1e2)$ for en genotype (hændelsespar) og multipliceres for at give den betingede sandsynlighed $P(R|e1e2)$ for at observere hele samlingen af læsninger. Ved hjælp af Bayes formel beregnes den posteriore sandsynlighed $P(e1e2|R)$ for hver diploid-genotype, og vinderen bestemmes.

DRAGEN for IDPE Dx anvender automatisk filtrering. Se [VCF-filannotationer om kimcellarbejdsgang på side 13](#) for yderligere oplysninger.

DRAGEN bestemmelse af somatiske varianter

DRAGEN Somatic Small Variant Caller tager tilknyttede og tilpassede DNA-læsninger som input og bestemmer SNV'er og indeler gennem lokal *de novo*-samling af haplotyper i det aktive område. For at aktivere DRAGEN Somatic Small Variant Caller skal du vælge et program til somatiske varianter.

Somatisk variationsbestemmelse bruges typisk til tumorprøver. Med denne arbejdsgang foretager DRAGEN ingen antagelser om ploidi, hvilket muliggør detektion af lavfrekvente alleler. For loci med dækning op til 100x i tumorprøven har DRAGEN en detektionsgrænse ved variationsallelfrekvenser på 5 %. Grænsen skalerer med stigende dybde pr. locus og halveres, hver gang dækningen fordobles ud over 100x. De referenceområder der kan bestemmes, identificeres først med tilstrækkelig tilpasningsdækning. Inden for disse referenceområder vil en scanning af de sorterede læsninger identificere aktive områder, som er centreret om samlingen af kolonner med evidens for en variant i tumorlæsningerne. De aktive områder er polstret med tilstrækkelig kontekst til at dække signifikant, ikke-reference-indhold i nærheden. Hvis der er evidens for indeler, vil de aktive områder få ekstra polstring.

Tilpassede læsninger klippes inden for hver aktivt område og samles i en De Bruijn-graf. Kanterne på de klippede læsninger vægtes af observationstal, med referencesekvensen som et skelet. Efter lidt oprydning og forenkling af grafen, ekstraheres alle source-to-sink-linjer som kandidat-haplotyper. Hver haplotype er Smith-Waterman tilpasset til referencegenomet for at identificere de varianter, den repræsenterer. For hvert læsning-haplotype-par estimeres sandsynligheden $P(r|H)$ for at observere læsningen ved brug af et par skjult Markov-model (HMM), når det antages at haplotypen er den sande startprøve.

For at bestemme scoren for påvisningsgrænsen for tumorer (TLOD), scanner DRAGEN Somatic Small Variant Caller først efter referenceposition for hver kandidat for somatisk hændelse samt referencehændelsen over det aktive område. Den betingede sandsynlighed $P(r|e)$ for at observere hver overlappende læsning, som den maksimale $P(r|H)$ for haplotyper, der understøtter denne hændelse. Disse kombineres for at give den betingede sandsynlighed $P(r|E)$ for en hændeshypotese, E, der involverer en blanding af referencen og kandidaten for somatisk allel over et område af mulige

allelfrekvenser og multipliceret for at give den betingede sandsynlighed $P(R|E)$ for at observere hele samlingen af læsninger. Derfra beregnes TLOD-scoren som evidens for, at en ALT-allel er til stede i tumorprøven ved en given locus.

DRAGEN for IDPE Dx anvender automatisk filtrering. Se [VCF-filannotationer om somatisk arbejdsgang på side 16](#) for yderligere oplysninger.

Opret en planlagt kørsel

Brug følgende trin for at oprette en kørsel i Illumina Run Manager enten på NextSeq 550Dx eller ved brug af en browser på en netværksforbundet computer. Brug en browser på en netværksforbundet computer, hvis du ønsker at importere prøvedata. Se Vejledning til softwaren Illumina Run Manager til NextSeq 550Dx (dokumentnr. 200025239) for instruktioner om adgang Illumina Run Manager via en netværksforbundet computer.

Der er to forskellige måder at oprette en ny, planlagt kørsel på:

- **Import Run** (Importér kørsel) – Brug et prøveark fra en eksisterende kørsel som skabelon for en ny kørsel. Se Vejledning til softwaren Illumina Run Manager til NextSeq 550Dx (dokumentnr. 200025239) for oplysninger om, hvordan du importerer en kørsel.
- **Create Run** (Opret kørsel) – Indtast kørselsparametre manuelt. Følgende anvisninger beskriver, hvordan man opretter en kørsel.

BEMÆRK De påkrævede indtastningsfelter i brugergrænsefladen er markeret med et asterisksymbol (*).

Program

1. Vælg **Create Run** (Opret kørsel) under fanen Planned (Planlagt) i skærmbilledet Runs (Kørsler).
2. Vælg DRAGEN for Illumina DNA Prep with Enrichment Dx-programmet, og vælg så **Next** (Næste).

Kørselsindstillinger

1. Indlæs et unikt kørselsnavn på skærmbilledet Run Settings (Kørselsindstillinger). Kørselsnavnet identificerer kørslen fra sekventering gennem analysen.
2. **[Valgfri]** Angiv en kørselsbeskrivelse for yderligere at identificere kørslen.
3. Vælg indeksadaptersæt, der bruges under klargøring af bibliotek.
4. Gennemgå læselængden, og rediger om nødvendigt. Læsning 1 og Læsning 2 har en standardværdi på 151 cyklusser. Indeks 1 og Indeks 2 har en fast værdi på 10 cyklusser, hvilket ikke kan ændres.
5. **[Valgfri]** Angiv et biblioteksror-ID.
6. Vælg **Next** (Næste).

Prøvedata

Prøvedata omfatter Sample ID (Prøve-ID), Well Position (Brøndplacering) (brøndplacering på indeksplassen) og Library Name (Biblioteksnavn). Når du bruger indeks A og B, omfatter Well Position (Brøndplaceringen) også en Plate identifier (Pladeidentifikator).

Der er to måder at indtaste prøvedata på:

- **Import Samples** (Importer prøver) – Brug en skabelonfil, der kan downloades på skærmen Sample Data (Prøvedata).
- **Manually** (Manuelt) – Indtast prøvedataene direkte i tabellen på skærmbilledet Sample Data (Prøvedata).

Importér prøver

Når man planlægger en sekvenskørsel via en browser på en netværksforbundet computer, kan man downloade en skabelonfil (*.CSV) fra skærmbilledet Sample Data (Prøvedata). Skabelonfilen kan ikke downloades, når du får adgang til Illumina Run Manager via NextSeq 550Dxoperativsystemet. Gør følgende for at indtaste prøvedata ved hjælp af funktionen Import Samples (Importér prøver).

BEMÆRK Udfør trinene for Run Settings (Kørselsindstillinger), før du fortsætter.

1. Vælg **Download Template** (Download skabelon) for at downloade en ikke-udfyldt CSV-fil.
2. Fra skabelonfilen skal du indtaste prøvedata og derefter gemme filen. Biblioteksnavnet er valgfrit.

BEMÆRK Når du bruger indeks A og B, skal dataene for kolonne B omfatte både plade- og brøndplacering (brøndplacering på indekspladen). Eksempel: A-A01, A-A02, A-A03.

3. Vælg **Import Samples** (Importér prøver), og gå til den skabelonfil, der indeholder prøvedata fra det forrige trin.
4. Vælg **Open** (Åbn), **Proceed** (Fortsæt) og derefter **Next** (Næste).

BEMÆRK Ændring af Sample ID (Prøve-ID), før du vælger Next (Næste), kan resultere i en fejl. Afslut opsætningen af kørslen, før du foretager ændringer for at undgå fejl.

Indtast prøver manuelt

Brug tabellen på skærmbilledet Sample Data (Prøvedata) til manuelt at angive prøveoplysningerne.

1. Angiv et unikt prøve-ID i feltet Sample ID (Prøve-ID).
2. Brug **Well Position** (Brøndplacering) (indeks A eller indeks B) eller **Plate – Well Position** (Plade – Brøndplacering) (Indeks A og B) til at vælge det indeks, der knyttes til prøverne. Felterne i7 Index (i7-Indeks), Index 1 (Indeks 1), i5 Index (i5-Indeks) og Index 2 (Indeks 2) udfyldes automatisk.
3. **[Valgfri]** Angiv et biblioteksnavn.

4. Tilføj rækker, og gentag trin 1–3 efter behov, indtil alle prøverne er føjet til tabellen. Du kan tilføje flere rækker ad gangen ved først at indtaste det antal rækker, der skal tilføjes, og derefter vælge ikonet +. Du kan også fjerne rækker ved at markere feltet ved siden af række nummeret og derefter klikke på papirkurvsikonet.
5. Vælg **Next** (Næste).

Analyseindstillinger

1. Vælg den ønskede analysearbejdsgang:
 - FASTQ-generering
 - FASTQ- og VCF-generering for en kimcellerarbejdsgang (manifestfil påkrævet)
 - FASTQ- og VCF-generering for en somatisk arbejdsgang (manifestfil påkrævet)
2. **[Valgfri] Generate ORA compressed FASTQs** (Generér ORA-komprimerede FASTQ-filer), er aktiveret som standard. FASTQ ORA-komprimering komprimerer tabsfri FASTQ-filer op til 5x sammenlignet med fastq.gz. Fjern markeringen **Generate ORA compressed FASTQs** (Generér ORA-komprimerede FASTQ-filer), hvis der foretrakkes ukomprimerede data (fastq.gz).
3. Der kræves en manifestfil til kimcellerarbejdsgange og somatiske arbejdsgange. Brug rullemenuen **Manifest File Selection** (Valg af manifestfil) for at vælge en manifestfil. Manifestfilen er en tabulatorseparatoreret BED (*.bed), der angiver navnene og placeringerne på de målrettede referenceområder. Se [Manifestfil på side 8](#) for yderligere oplysninger.
4. **[Valgfri]** For somatiske arbejdsprocesser skal du bruge rullemenuen **Noise File Selection** (Valg af støjfil) til at vælge en fil for systematisk støj. Der kan angives en BED-fil (*.bed.gz) med centerspecifikt støjniveau til at frasortere systematisk støj. Se [Støjfiltrering \(Valgfri\) på side 9](#) for yderligere oplysninger.
5. Vælg **Next** (Næste).

Kørsels- oversigt

1. På skærbilledet Review (Oversigt) skal du gennemgå oplysningerne for Run Settings (Kørselsindstillinger), Sample Data (Prøvedata) og Analysis Settings (Analyseindstillinger).
2. Vælg **Save** (Gem).
Kørslen gemmes under fanen Planned (Planlagt) på skærmen Runs (Kørsler).

Indstillinger

For at få vist eller foretage ændringer i DRAGEN for IDPE Dx programindstillinger skal du først vælge ikonet Applications (Programmer) på hovedskærmen. Vælg derefter det program, du vil have vist eller ændre. Der kræves en administrator-konto for at ændre indstillingerne.

Konfiguration

Konfigurationsskærmen viser følgende programindstillinger:

- **Library Prep Kits** (Biblioteksklarkøringsæt) – Viser standardbiblioteksklarkøringssettet for programmet. Denne indstilling kan ikke ændres.
- **Index Adapter Kits** (Indeksadaptersæt) – Viser standard indeksadaptersæt for programmet. Denne indstilling kan ikke ændres.
- **Read lengths** (Læsningslængder) – Læsningslængder er indstillet til 151 på programmet som standard, men kan ændres under oprettelse af kørslen.
- **Manifest and Noise Files** (Manifest- og støjfiler) – Upload, og ændr indstillinger for manifest- og støjfiler.
 - Vælg **Upload File** (Upload fil) for at uploade fil til brug i analyse.
 - Vælg alternativknappen **Default** (Standard) for at indstille filen som standard manifest- eller støjfil under oprettelse af kørslen, når programmet er valgt.
 - Vælg afkrydsningsfeltet **Enabled** (Aktiveret) for at indstille filen til at blive vist i rullemenuen under oprettelse af kørslen.

Rettigheder

Brug afkrydsningsfelterne på skærmen Permissions (Godkendelser) to at administrere brugeradgang til appen.

Manifestfil

Når du bruger DRAGEN for IDPE Dx, er en manifestfil påkrævet input for følgende arbejdsgange:

- FASTQ- og VCF-generering for en kimcellearbejdsgang
- FASTQ- og VCF-generering for en somatisk arbejdsgang

Manifestfilen er en tabulatorsepareret tekstfil, der gør brug af BED-formatet (*.bed), der angiver navnene og placeringerne på de målrettede referenceområder. Den centrale del af manifestfilen er afsnittet Regions (Områder), og den skal indeholde følgende datakolonner:

Kolonne	Beskrivelse
Navn	Unikt brugerangivet navn for målet
Kromosom	Kromosomplacering (f.eks. chr10, chr5 osv.)
Start	1-baseret indeks for målets startposition
Stop	1-baseret indeks for målets stopposition
Probelængde opstrøms	Længden af proben opstrøms. For appen DRAGEN for IDPE Dx skal dette indstilles til 0.
Probelængde nedstrøms	Længden af proben nedstrøms. For appen DRAGEN for IDPE Dx skal dette indstilles til 0.

BEMÆRK Der kræves et gyldigt manifestfilformat til analyse. DRAGEN stopper analysen, hvis manifestfilen er ugyldig.

Støjfiltrering (Valgfri)

Filtrering af systematisk støj er tilgængeligt til bestemmelse af somatiske variationer og kan bruges til at reducere falske positive bestemmelser ved at tage højde for stedspecifik støj. Den systematiske støjfil genereres ved først at indsamle ca. 50 normale prøver (helst specifikke for panelet, biblioteksklargøring og sequenceren), og derefter divideres summen af allefrekvenser, under 30 % på hvert sted med tilstrækkelig dækning, med det samlede antal prøver (alle frekvenser over 30 % antages at være kimcellevariationer og ikke støj). Når støjværdierne er genereret, vil somatiske varianter, der er detekteret på dette sted, blive filtrerede.

Filteret kan bruges i tilstanden Tumor-Normal, men er især praktisk anvendeligt i kørslerne Tumor-Only (Kun tumor), hvor en matchet normal ikke er tilgængelig. Den systematiske støjfil skal bruge en BED-fil, der har et filtypenavn (* .bed.gz), der skal indeholde fire kolonner: Chromosome, Start, End (Kromosom, Start, Slut) og stedspecifikke støjniveauer for hver række. Filtrering af systematisk støj er valgfri.

Analyseoutput

Kørsler, der aktuelt er i gang, vises under fanen Active (Aktiv). Fuldførte kørsler vises i fanen Completed (Fuldført). DRAGEN for IDPE Dx opretter en analysemappe med et unikt navn for hver analyse, som er adskilt fra mappen med sekventeringsdata. Analysemappen gemmer følgende oplysninger:

- Manifestfil brugt
- Softwareversion
- Prøve-ID'er
- Tilpassede læsninger i alt
- Procentdel af tilpassede læsninger pr. prøve
- Antal bestemte SNV'er pr. prøve
- Antal bestemte indeler pr. prøve
- Dækningsstatistik

Analyseoutputfiler

Placeringen af analysemappen er angivet af indstillingen for External Storage for Analysis Results (Ekstern opbevaring af analyseresultater). Se Vejledning til softwaren Illumina Run Manager til NextSeq 550Dx (dokumentnr. 200025239) for yderligere oplysninger om indstillingen External Storage for Analysis Results (Ekstern opbevaring af analyseresultater).

På skærmbilledet Run Details (Kørselsoplysninger) viser feltet External Location (Ekstern placering) stien til sekventeringsdata. Navnet på den unikke analysemappe findes i feltet Analysis Output (Analyseoutput) på skærmbilledet Run Details (Kørselsoplysninger). De eksakte, genererede filer vil afhænge af, hvilken analysearbejdsgang der bruges. Programmet vil generere følgende analyseoutputfiler.

BEMÆRK Hvis der opstår en fejl med begrænsning af filstiens maksimale længde, når du tilgår analyseoutputfiler, kan du prøve at flytte filen til en kortere stiplacering eller bruge en anden metode til at åbne filen.

Outputfil	Beskrivelse
Oversigtsrapport over varianter (*.pdf)	Indeholder en oversigt over filoplysninger, softwareversioner, eksempeloplysninger, statistik på læsningsniveau og SNV, indsættelser, sletninger og dækningsoversigter. Det er kun kimcellerarbejdsgange og somatiske arbejdsgange, der genererer en rapport over varianter.
FASTQ (*.fastq.gz eller *.fastq.ora)	Midlertidige filer med kvalitetsscorede basebestemmelser. FASTQ-filer er det primære input til tilpasningstrinnet. Når ORA-komprimering er valgt, bruges filtypenavnet *.fastq.ora.
Tilpasnings-BAM-filer (*.bam)	Indeholder tilpassede læsninger for en given prøve.
Genome-VCF-filer (*.gvcf.gz)	Indeholder genotype for hver placering, uanset om denne er bestemt som en variant eller bestemt som en reference.
VCF-filer (*.vcf.gz)	Indeholder varianter bestemt ved hver placering.
Kørselsmålingsrapport (*.csv)	Indeholder kvalitetsmålinger om kørslen, herunder samlet resultat uden indeksering og Q30-score.

FASTQ-filer

FASTQ (*.fastq.gz, *.fastq.ora) er et tekstbaseret filformat, der indeholder basebestemmelser og kvalitetsværdier for hver læsning. Hver fil indeholder følgende oplysninger:

- Prøveidentifikatoren
- Sekvensen
- Phred-kvalitetsscorerne i et ASCII + 33-kodet format

Prøveidentifikatoren er formateret som følger:

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y
ReadNum:FilterFlag:0:SampleNumber
```

```

Eksempel :
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAAA9#:<#<;<<<????#=#

```

BAM-filer

En BAM-fil (*.bam) er en komprimeret, binær version af en SAM-fil (Sequence Alignment Map), der bruges til at repræsentere tilpassede sekvenser på op til 128 Mb. BAM-filer bruger filnavnformatet `SampleName_S#.bam`. # er prøvenummeret ud fra den rækkefølge, som prøverne er angivet i for kørslen. I multinode-tilstanden er S# indstillet til S1, uanset prøverækkefølgen.

BAM-filer indeholder en hovedsektion og en tilpasningssektion:

- **Header** (Hoved) – Indeholder oplysninger om hele filen, såsom prøvenavn, prøvelængde og tilpasningsmetode. Tilpasninger i tilpasningssektionen er forbundet med specifikke oplysninger i hovedsektionen.
- **Alignments** (Tilpasninger) – Indeholder læsningsnavn, læsningssekvens, læsningskvalitet, tilpasningsoplysninger og brugerdefinerede tags. Læsningsnavnet inkluderer kromosomet, startkoordinaten, tilpasningskvaliteten og den matchede beskrivelsesstreng.

Tilpasningssektionen inkluderer følgende oplysninger for hver læsning eller hvert læsningspar:

- AS: Paired-end-tilpasningskvalitet.
- RG: Læsningsgruppe, hvilket indikerer antallet af læsninger for en specifik prøve.
- BC: Stregkode-tag, hvilket angiver det demultiplexerede prøve-ID, der er forbundet med læsningen.
- SM: Single-end-tilpasningskvalitet.
- XC: Matchbeskrivelsesstreng.
- XN: Amplicons navnetag, der registrerer det amplicon-ID, der er forbundet med læsningen.

BAM-indeksfiler (*.bam.bai) giver et indeks over den tilhørende BAM-fil.

VCF-filer

Variant Call Format-filerne (*.vcf) indeholder oplysninger om varianter fundet ved specifikke placeringer i et referencegenom.

VCF-filhovedet inkluderer VCF-filformatets version, variantbestemmelsesprogrammets version og en liste over de annoteringer, der er brugt i resten af filen. VCF-hovedet inkluderer også referencegenomfilen og BAM-filen. Den sidste linje i hovedet indeholder kolonneoverskrifter for datalinjerne. Hver af VCF-filens datalinjer indeholder oplysninger om en enkelt variant.

Tabel 1 VCF-filens overskrifter

Overskrift	Beskrivelse
CHROM	Referencegenomets kromosom. Kromosomerne vises i samme rækkefølge som FASTA-referencefilen.
POS	Variantens enkelt-base-placering i referencekromosomet. For SNV'er (enkeltnukleotidvarianter) er denne placering referencebasen med varianten. For indeler er denne placering referencebasen umiddelbart før varianten.
ID	rs-tallet (reference-SNP) for den SNP, der er indhentet fra <code>dbSNP.txt</code> , hvis relevant. Hvis der findes flere rs-tal på denne placering, er listen semikolonsepareret. Hvis der ikke findes en dbSNP-post på denne placering, bruges der en markør for den manglende værdi ('.').
REF	Referencegenotypen. For eksempel er en sletning af et enkelt T repræsenteret som reference-TT og alternativ- T. En A til T-enkeltnukleotidvariant er repræsenteret som reference-A og alternativ-T.
ALT	De alleler som er forskellige fra referencelæsningen. En indsættelse af et enkelt T er for eksempel repræsenteret som reference-A og alternativ-AT. En A til T-enkeltnukleotidvariant er repræsenteret som reference-A og alternativ-T.
QUAL	En Phred-skala-kvalitetsscore tildelt af variantbestemmelsesprogrammet. Højere scorer indikerer større tillid til varianten og lavere sandsynlighed for fejl. For en kvalitetsscore på Q er den estimerede sandsynlighed for en fejl $10^{-(Q/10)}$. For eksempel har sættet med Q30-bestemmelser en fejlrate på 0,1 %. Mange variantbestemmelsesprogrammer tildeler kvalitetsscorer baseret på deres statistiske modeller, som er høje i forhold til den observerede fejlrate.

Tabel 2 VCF-filannotationer om kimcellerarbejdsgang

Overskrift	Beskrivelse
FILTER	<p>Hvis alle filtrene passeres, vil der stå PASS (BESTÅET) i filterkolonnen. Mulige FILTER-poster omfatter:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL – Bruges hvis SNP-variantens QUAL-score ikke opfylder grænseværdien • DRAGENIndelHardQUAL – Bruges hvis indelvariantens QUAL-score ikke opfylder grænseværdien • LowDepth – Filtreret af centret, fordi dybden af dækningen ikke opfylder grænseværdien • LowGQ – Filtreret af centret, fordi genotypekvaliteten ikke opfylder grænseværdien • PloidyConflict – Genotypebestemmelse fra variantbestemmelsesprogram er ikke i overensstemmelse med kromosomploidi • base_quality – Filtreret af centret, fordi den gennemsnitlige basekvalitet af alt-læsningerne ved denne locus ikke opfylder grænseværdien • filtered_reads – Filtreret af centret, fordi en for stor andel af læsningerne er blevet frasorteret • fragment_length – Filtreret af centret, fordi den absolutte forskel på den gennemsnitlige fragmentlængde af alt-læsninger og den gennemsnitlige fragmentlængde af ref-læsninger ved denne locus er over grænseværdien • low_depth – Filtreret af centret, fordi læsningsdybden er for lav • low_frac_info_reads – Filtreret af centret, fordi fraktionen af informative læsninger er under grænseværdien • low_normal_depth – Filtreret af centret, fordi den normale prøvelæsningsdybde er for lav • long_indel – Filtreret af centret, fordi indellængden er for lang • mapping_quality – Filtreret af centret, fordi den gennemsnitlige basekvalitet af alt-læsningerne ved denne locus ikke opfylder grænseværdien • multiallelic – Filtreret af centret, fordi mere end to alt-alleler passerer tumor-LOD • non_homref_normal – Filtreret af centret, fordi den normale prøvegenotype ikke er homozygot-reference • no_reliable_supporting_read – Filtreret af centret, fordi der ikke findes nogle pålidelige, understøttende, somatiske læsninger • panel_of_normals – Set i mindst en prøve i panelet med normale vcf • read_position – Filtreret af centret, fordi medianen af afstande mellem start/slut af læsningen, og denne locus er under grænseværdien • RMxNRepeatRegion – Filtreret af centret, fordi alle eller en del af variantallelen er en gentagelse af referencen • strand_artifact – Filtreret af centret på grund af alvorlig streng-bias • str_contraction – Filtreret af centret på grund af formodet PCR-fejl, hvor alt-allelen er en gentagelsesenhed under referencen

Overskrift	Beskrivelse
FILTER	<ul style="list-style-type: none"> • too_few_supporting_reads – Filtreret af centret, fordi der er for få understøttende læsninger i tumorprøven • weak_evidence – Somatisk variantscore opfylder ikke grænseværdien
INFO	<p>Mulige INFO-poster omfatter:</p> <ul style="list-style-type: none"> • AC – Alleltal i genotyper for hver ALT-allel i samme rækkefølge som angivet. • AF – Allelfrekvens for hver ALT-allel i samme rækkefølge som angivet. • AN – Samlet antal alleler i bestemte genotyper. • DB – dbSNP-medlemskab. • FS – P-værdi på Phred-skala ved brug af Fishers eksakte test til påvisning af bias af strenge. • QD – Variantkonfidens/-kvalitet efter dybde. • R2_5P_bias – Score baseret på bias af mate og afstand fra 5 prim-ende. • SOR – Symmetrisk odds-forhold for 2x2 kontingenstabel til påvisning af bias af strenge. • DP – Omtrentlig læsningsdybde (informativ og ikke-informativ). Nogle læsninger kan være blevet filtreret på mapq osv. • END – Intervallets stopposition. • FractionInformativeReads – Fraktionen af informative læsninger ud af det samlede antal læsninger. • MQ – RMS-tilknytningskvalitet. • MQRankSum – Z-score fra Wilcoxons rangsumstest af kvaliteten af Alt- vs. Ref-læsningstilknytninger. • ReadPosRankSum – Z-score fra Wilcoxons rangsumstest af bias af Alt vs. Ref-læsningsplaceringer. • SOMATIC – Mindst en variant på denne placering er somatisk.

Overskrift	Beskrivelse
FORMAT	<p>Formatkolonnen har kolonseparerede felter. For eksempel GT:GQ. Mulige tilgængelige felter:</p> <ul style="list-style-type: none"> • AD – Alleldybder (tæller kun informative læsninger ud af det samlede antal læsninger) for ref- og alt-allelerne i den rækkefølge, de er angivet. • AF – Allelfraktioner for alt-alleler i den rækkefølge, de er angivet. • DP – Omtrentlig læsningsdybde (læsninger med MQ = 255 eller med dårlige mates filtreres). • F1R2 – Antal læsninger i F1R2-parretning, der understøtter hver allel. • F2R1 – Antal læsninger i F2R1-parretning, der understøtter hver allel. • GT – Genotype. 0 svarer til referencebasen, 1 svarer til den første post i ALT-kolonnen osv. Skråstregen (/) indikerer, at der ikke er tilgængelige oplysninger om inddeling i faser. • MB – Statistik pr. prøvekomponent til påvisning af mate-bias. • PS – ID-oplysninger for fysisk faseinddeling, hvor hvert unikt ID i en given prøve (men ikke på tværs af prøver) forbinder poster inden for en faseinddelingsgruppe. • SB – Statistik pr. prøvekomponent, som består af Fishers eksakte test til påvisning af bias af strenge. • SQ – Somatisk kvalitet.
SAMPLE (PRØVE)	Prøvekolonnen giver de værdier, der er angivet i kolonnen FORMAT.

Tabel 3 VCF-filannotationer om somatisk arbejdsgang

Overskrift	Beskrivelse
FILTER	<p>Hvis alle filtrene passerer, vil der stå PASS (BESTÅET) i filterkolonnen. Mulige FILTER-poster omfatter:</p> <ul style="list-style-type: none"> • base_quality – Filtreret af centret, fordi den gennemsnitlige basekvalitet af alt-læsningerne ved denne locus ikke opfylder grænseværdien • filtered_reads – Filtreret af centret, fordi en for stor fraktion af læsningerne er blevet frasorteret • fragment_length – Filtreret af centret, fordi den absolutte forskel på den gennemsnitlige fragmentlængde af alt-læsninger og den gennemsnitlige fragmentlængde af ref-læsninger ved denne locus er over grænseværdien • low_depth – Filtreret af centret, fordi læsningsdybden er for lav • low_frac_info_reads – Filtreret af centret, fordi fraktionen af informative læsninger er under grænseværdien • low_normal_depth – Filtreret af centret, fordi den normale prøvelæsningsdybde er for lav • long_indel – Filtreret af centret, fordi indellængden er for lang • mapping_quality – Filtreret af centret, fordi den gennemsnitlige basekvalitet af alt-læsningerne ved denne locus ikke opfylder grænseværdien • multiallelic – Filtreret af centret, fordi mere end to alt-alleler passerer tumor-LOD • non_homref_normal – Filtreret af centret, fordi den normale prøvegenotype ikke er homozygot-reference • no_reliable_supporting_read – Filtreret af centret, fordi der ikke findes nogle pålidelige, understøttende, somatiske læsninger • panel_of_normals – Set i mindst en prøve i panelet med normale vcf • read_position – Filtreret af centret, fordi medianen af afstande mellem start/slut af læsningen, og denne locus er under grænseværdien • RMxNRepeatRegion – Filtreret af centret, fordi alle eller en del af variantallelen er en gentagelse af referencen • strand_artifact – Filtreret af centret på grund af alvorlig streng-bias • str_contraction – Filtreret af centret på grund af formodet PCR-fejl, hvor alt-allelen er en gentagelsesenhed under referencen • too_few_supporting_reads – Filtreret af centret, fordi der er for få understøttende læsninger i tumorprøven • weak_evidence – Somatisk variantscore opfylder ikke grænseværdien • systematic_noise – Filtreret af centret baseret på evidens for systematisk støj i normaler

Overskrift	Beskrivelse
INFO	<p>Mulige INFO-poster omfatter:</p> <ul style="list-style-type: none"> • DP – Omtrentlig læsningsdybde (informativ og ikke-informativ). Nogle læsninger kan være blevet filtreret på mapq osv. • END – Intervallets stopposition. • FractionInformativeReads – Fraktionen af informative læsninger ud af det samlede antal læsninger. • MQ – RMS-tilknytningskvalitet. • MQRankSum – Z-score fra Wilcoxons rangsumstest af kvaliteten af Alt- vs. Ref-læsningstilknytninger. • ReadPosRankSum – Z-score fra Wilcoxons rangsumstest af bias af Alt vs. Ref-læsningsplaceringer. • AQ – Score for systematisk støj. • hotspot – Kendt somatisk sted, brugt til at øge tilliden til bestemmelsen. • SOMATIC – Mindst en variant på denne placering er somatisk.
FORMAT	<p>Formatkolonnen har kolonseparerede felter. For eksempel GT:GQ. Mulige tilgængelige felter:</p> <ul style="list-style-type: none"> • AD – Allellybder (tæller kun informative læsninger ud af det samlede antal læsninger) for ref- og alt-allelerne i den rækkefølge, de er angivet. • AF – Allelfraktioner for alt-alleler i den rækkefølge, de er angivet. • DP – Omtrentlig læsningsdybde (læsninger med MQ = 255 eller med dårlige mates filtreres). • F1R2 – Antal læsninger i F1R2-parretning, der understøtter hver allel. • F2R1 – Antal læsninger i F2R1-parretning, der understøtter hver allel. • GP – Posteriore sandsynligheder for genotyper ifølge Phred-skalaen, som defineret i VCF-specifikationen. • GQ – Genotypekvalitet. • GT – Genotype. 0 svarer til referencebasen, 1 svarer til den første post i ALT-kolonnen osv. Skråstregen (/) indikerer, at der ikke er tilgængelige oplysninger om inddeling i faser. • MB – Statistik pr. prøvekomponent til påvisning af mate-bias. • PL – Normaliseret sandsynlighed for genotyper ifølge Phred-skalaen, som defineret i VCF-specifikationen. • PRI – Tidligere sandsynlighed for genotyper ifølge Phred-skalaen. • PS – ID-oplysninger for fysisk faseinddeling, hvor hvert unikt ID i en given prøve (men ikke på tværs af prøver) forbinder poster inden for en faseinddelingsgruppe. • SB – Statistik pr. prøvekomponent, som består af Fishers eksakte test til påvisning af bias af strenge. • SQ – Somatisk kvalitet.

Overskrift	Beskrivelse
SAMPLE (PRØVE)	Prøvekolonnen giver de værdier, der er angivet i kolonnen FORMAT.

Genome VCF-filer

Genome VCF-filer (*.gvcf.gz) følger et sæt konventioner til repræsentation af alle steder i genomet i et rimeligt kompakt format. gVCF-filerne inkluderer alle steder i interesseområdet i en enkelt fil for hver prøve. gVCF-filen viser manglende bestemmelser ved placeringer, der ikke passerer alle filtrene. Et genotype-tag (GT) med ./ indikerer en manglende bestemmelse.

Genindsættelse i analysekø

Du kan sætte en analyse, som er blevet stoppet, i kø igen, hvis analysen mislykkedes, eller hvis du vil analysere en kørsel igen med andre indstillinger. Udfør følgende trin for at sætte analysen i kø igen:

- Vælg fanen Completed (Fuldført) på skærbilledet Run (Kørsel), og vælg derefter det kørselsnavn, der skal analyseres igen.
Hvis der tidligere blev udført Requeue Analysis (Genindsæt i analysekø), skal du vælge kørselsnavnet for Parent Run (Overordnet kørsel).
- Fra skærmen Run Details (Kørselsoplysninger), efter Sequencing Information (Sekventeringsoplysninger), skal du vælge **Requeue Analysis** (Genindsæt i analysekø).
- Vælg en mulighed:
 - Requeue analysis with no changes (Sådan genindsættes i analysekø)
 - Edit run settings and requeue analysis (Rediger kørselsindstillinger, og genindsæt analyse i kø)
 - Requeue analysis with a different application (Genindsæt analyse i køen med et andet program)
- Bekræft, at placeringen, hvor sekventeringsdataene aktuelt befinder sig, findes i feltet **Sequencing data file path** (Sti til sekventeringsdatafilen).

BEMÆRK Stien til sekventeringsdataene skal stemme overens med stien i indstillingen External Storage for Analysis Results (Ekstern opbevaring af analyseresultater). Se Vejledning til softwaren Illumina Run Manager til NextSeq 550Dx (dokumentnr. 200025239) for oplysninger om ændring af den eksterne opbevaringssti.

- Reanalysis Reason (årsag til at der skal analyseres).
- Vælg **Requeue Analysis** (Genindsæt i analysekø).
- Rediger de ønskede ændringer i indstillingerne for Run Settings (Kørselsindstillinger) Sample Data (Prøvedata) og Analysis Settings (Analyseindstillinger).

8. Vælg **Save** (Gem). Analysen starter med de aktuelle analyseparametre.

Teknisk hjælp

Kontakt Illumina teknisk support for at få teknisk hjælp.

Websted: www.illumina.com

E-mail: techsupport@illumina.com

Sikkerhedsdatablade (SDS'er) – Kan findes på Illuminas websted på support.illumina.com/sds.html.

Produktdokumentation – Kan downloades på support.illumina.com.



Illumina
5200 Illumina Way
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (uden for Nordamerika)
techsupport@illumina.com
www.illumina.com



Australsk sponsor

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

TIL IN VITRO DIAGNOSTISK BRUG.

© 2023 Illumina, Inc. Alle rettigheder forbeholdes.

illumina[®]