

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

Brukerveiledning for applikasjon

ILLUMINA-PROPRIETÆR

Dokumentnr. 200025238 v00

Februar 2023

TIL IN VITRO-DIAGNOSTISK BRUK.

Dette dokumentet og dets innhold er opphavsrettslig beskyttet for Illumina, Inc. og dets tilknyttede selskaper («Illumina»), og er ment utelukkende for kontraktbruk av kunden i forbindelse med bruk av produktet (produktene) beskrevet her, og for intet annet formål. Dette dokumentet og dets innhold skal ikke brukes eller distribueres til andre formål og/eller på annen måte kommuniseres, fremlegges eller reproduseres på noen måte uten forutgående, skriftlig samtykke fra Illumina. Illumina overfører ikke noen lisens under sitt patent, varemerke, opphavsrett eller sedvanerett eller lignende rettigheter til tredjeparter gjennom dette dokumentet.

Instruksjonene i dette dokumentet skal følges nøyaktig og kun av kvalifisert og tilfredsstillende utdannet personell for å sikre riktig og sikker bruk av produktet (produktene) som er beskrevet i dette dokumentet. Alt innhold i dette dokumentet skal leses fullt ut og være forstått før produktet (produktene) brukes.

HVIS DET UNNLATES Å LESE FULLSTENDIG OG UTTRYKkelig FØLGE ALLE INSTRUKSJONENE I DETTE DOKUMENTET, KAN DET FØRE TIL SKADE PÅ PRODUKTET (PRODUKTENE), SKADE PÅ PERSONER, INKLUDERT BRUKERE ELLER ANDRE, OG SKADE PÅ ANNEN EIENDOM, OG DETTE VIL UGYLDIGGJØRE EVENTUELL GARANTI SOM GJELDER FOR PRODUKTET (PRODUKTENE).

ILLUMINA PÅTAR SEG IKKE ANSVAR SOM FØLGE AV FEIL BRUK AV PRODUKTET (PRODUKTENE) SOM ER BESKREVET I DETTE DOKUMENTET (INKLUDERT DELER AV DETTE ELLER PROGRAMVARE).

© 2023 Illumina, Inc. Alle rettigheter forbeholdt.

Alle varemerker tilhører Illumina, Inc. eller deres respektive eiere. For spesifikk informasjon om varemerker, se www.illumina.com/company/legal.html.

Revisjonshistorikk

Dokument	Dato	Beskrivelse av endring
200025238 v00	Februar 2023	Første versjon.

Innholdsfortegnelse

Revisjonshistorikk	iii
Oversikt	1
Analysemetoder	1
Opprett en planlagt kjøring	5
Innstillinger	7
Manifestfil	8
Støyfiltrering (valgfritt)	9
Analyseutdataer	9
FASTQ-filer	10
BAM-filer	11
VCF-filer	11
Requeue Analysis (Sett analyse tilbake i kø)	17
Teknisk hjelp	18

Oversikt

DRAGEN for Illumina DNA Prep with Enrichment Dx-applikasjonen (DRAGEN for IDPE Dx) brukes til å planlegge og utføre sekundæranalyse av IDPE Dx-biblioteker generert for sekvensering på NextSeq 550Dx.

DRAGEN for IDPE Dx støtter sekvensering til analyse når den brukes med Illumina DNA Prep with Enrichment Dx bibliotekklargjøringen, NextSeq 550Dx, og Illumina DRAGEN Server for NextSeq 550Dx.

Analysemetoder

DRAGEN for IDPE Dx utfører demultipleksing, FASTQ-generering, kartlesing, justering til et referansegenom og små variantbetegnelse avhengig av de valgte arbeidsflytene:

- FASTQ-generering
- Germline FASTQ- og VCF-generering
- Somatisk FASTQ- og VCF-generering

MERK ORA-komprimering er tilgjengelig for bruk med alle tre arbeidsflyter. DRAGEN ORA Compression er en fullstendig tapsfri komprimeringsprogramvare som oppretter en fil med en Original Read Archive (*.ora)-forlengelse. Ora-formatet er et referansebasert komprimeringsformat for FASTQ-filer og er utformet for svært rask komprimering/dekomprimering og høyt komprimeringsforhold.

FASTQ Generasjon

De sammensatte sekvensene skrives til FASTQ-filer per prøve. FASTQ-filer er tekstfiler som inneholder sekvenseringsdata og kvalitetspoeng for kun én prøve. For hver prøve genereres separate FASTQ-filer per strømningscellefelt, per sekvensavlesning. Navnet på prøven som spesifisert under kjøringssoppsett er inkludert i FASTQ-filnavnet. FASTQ-filer er primære inndata for innretting. Det første trinnet i FASTQ-generering er demultipleksing. Demultipleksing tilordner klynger som sender filter til en prøve ved å sammenligne hver indekslesesekvens med indekssekvensene spesifisert for kjøringen. Ingen kvalitetsverdier blir vurdert i dette trinnet. Indeksavlesninger blir identifisert ved hjelp av følgende trinn:

- Prøver blir nummerert med start fra 1 basert på rekkefølgen de er listet opp for kjøringen.
- Prøvenummer 0 er reservert for klynger som ikke ble tilordnet en prøve.
- Klynger blir tilordnet en prøve når indekssekvensen samsvarer nøyaktig eller når det er opptil et enkelt misforhold per indeksavlesning.

Programvaren inkluderer ORA-komprimering for å komprimere FASTQ-filer. Dette formatet kan eventuelt aktiveres. Når du bruker ORA (*.ora)-formatet, bevares md5-sjekksummen av FASTQ-innholdet etter en komprimerings- og dekomprimeringssyklus for å sikre en tapsfri komprimering.

DNA-kartlegging og justering

Etter FASTQ-generering, kartlegges og justeres avlesningene til et referansegenom. Den første fasen av kartleggingen er å generere frø fra avlesningen, og deretter se etter eksakte treff i referansegenomet. Disse resultatene foredles deretter ved å kjøre fulle Smith-Waterman-justeringer på stedene med høyest tetthet av frøtreff. Denne veldokumenterte algoritmen fungerer ved å sammenligne hver posisjon av lesingen med alle kandidatposisjonene til referansen. Disse sammenligningene tilsvarer en matrise av potensielle justeringer mellom avlesning og referanse. For hver av disse kandidatjusteringsposisjonene genererer Smith-Waterman scoringer som brukes til å evaluere om den beste justeringen som passerer gjennom den matrisecellen når den ved en nukleotidmatch eller mismatch (diagonal bevegelse), en delesjon (horisontal bevegelse) eller en innsetting (vertikal bevegelse). En match mellom avlesning og referanse gir en bonus på poengsummen, og en mismatch eller indel gir en straff. Den generelle banen med høyest poengsum gjennom matrisen er den valgte justeringen. Algoritmen er maskinvareakselerert på DRAGEN-feltprogrammerbare gate array (FPGA)-kortene. Referansegenomet som brukes i appen er opprettet fra UCSC hg19 FASTA med DRAGEN-alternativet for å opprette en liftover-basert alt-aware hashtabell.

DRAGEN Germline Variant Calling

DRAGEN Kimlinje liten variantbetegnelse tar kartlagte og justerte DNA-avlesninger som input og betegner enkeltnukleotidpolymorfismer (SNP-er) og innsetting eller sletting (indeler) gjennom en kombinasjon av kolonnevis deteksjon og lokal *de novo*-sammenstilling av haplotyper. For å aktivere DRAGEN Kimlinje liten variantbetegnelse, velg arbeidsflyten for kimlinjevariant.

Kimlinjevariantbetegnelse brukes vanligvis for kimlinjeprøver der ploidi er kjent for å være to. Betegnelige referanseregioner identifiseres først med tilstrekkelig innrettingsdekning. Innenfor disse referanseområdene identifiserer en rask skanning av de sorterte avlesningene aktive regioner, som er sentrert på pileup-kolonner med bevis på en variant. De aktive områdene er polstret med nok kontekst til å dekke betydelig, ikke-referanseinnhold i nærheten. Hvis det er bevis på indels, får de aktive regionene ekstra polstring.

Justerte avlesninger klippes innenfor hver aktive region og settes sammen til en De Bruijn-graf. Kantene på de klippede avlesningene er vektet av observasjonstillinger, med referansesekvensen som en ryggrad. Etter litt grafrydding og forenkling, blir alle kilde-til-synk-baner trukket ut som kandidathaplotyper. Hver haplotype er Smith-Waterman justert til referansegenomet for å identifisere variantene den representerer. Dette settet med hendelser kan forsterkes med en posisjonsbasert deteksjon. For hvert avlesningshaplotypepar estimeres sannsynligheten $P(r|H)$ for å observere avlesningen, forutsatt at haplotypen er den sanne startprøven, ved å bruke en par skjult Markov-modell (HMM).

Skanning etter referanseposisjon over den aktive regionen, er kandidatgenotyper dannet fra diploide kombinasjoner av varianthendelser (SNP eller indels). For hver hendelse (inkludert referanse) estimeres den betingede sannsynligheten $P(r|e)$ for å observere hver overlappende avlesning som maksimal $P(r|H)$ for haplotyper som støtter hendelsen. Disse kombineres til den betingede sannsynligheten $P(r|e1e2)$ for en genotype (hendelsespar) og multipliseres for å gi den betingede sannsynligheten $P(R|e1e2)$ for å observere hele den avleste pileupen. Ved å bruke Bayes' formel beregnes den bakre sannsynligheten $P(e1e2|R)$ for hver diploid genotype, og vinneren betegnes.

DRAGEN for IDPE Dx bruker automatisk filtrering. Se [VCF-filmerknader for kimlinjearbeidsflyt på side 13](#) for mer informasjon.

DRAGEN Somatic Variant Calling

DRAGEN Somatisk liten variantbetegnelse tar kartlagte og justerte DNA-lesninger som input og betegner SNVer og indeler gjennom lokal *de novo*-sammenstilling av haplotyper i en aktiv region. For å aktivere DRAGEN Somatisk liten variantbetegnelse, velg et somatisk variantprogram.

Somatisk variantbetegnelse brukes vanligvis for tumorprøver. Med denne arbeidsflyten gjør ikke DRAGEN noen ploidiatangelser, noe som muliggjør deteksjon av lavfrekvente alleler. For loci med dekning opptil 100x i tumorprøven har DRAGEN en deteksjonsterskel ved variantallelfrekvenser på 5 %. Grensen skaleres med økende dybde på en per-locus basis og halveres hver gang dekningen doubles utover 100x. Betegnelige referanseregioner identifiseres først med tilstrekkelig innrettingsdekning. Innenfor disse referanseregionene identifiserer en skanning av de sorterte avlesningene aktive regioner, som er sentrert på pileup-søyler med bevis på en variant i tumor-avlesningene. De aktive områdene er polstret med nok kontekst til å dekke betydelig, ikke-referanseinnhold i nærheten. Hvis det er bevis på indels, får de aktive regionene ekstra polstring.

Justerte avlesninger klippes innenfor hver aktive region og settes sammen til en De Bruijn-graf. Kantene på de klippede avlesningene er vektet av observasjonstillinger, med referansesekvensen som en ryggrad. Etter litt grafrydding og forenkling, blir alle kilde-til-synk-baner trukket ut som kandidathaplotyper. Hver haplotype er Smith-Waterman justert til referansegenomet for å identifisere variantene den representerer. For hvert lese-haplotype-par estimeres sannsynligheten $P(r|H)$ for å observere avlesningen ved å bruke en par skjult Markov-modell (HMM) forutsatt at haplotypen er den sanne startprøven.

For å bestemme scoren for tumordeteksjonsgrense (TLOD), skanner DRAGEN Somatisk liten variantbetegnelse først etter referanseposisjon for hver kandidats somatiske hendelse samt referansehendelsen over den aktive regionen. Den betingede sannsynligheten $P(r|e)$ av å observere hver overlappende avlesning som maksimal $P(r|H)$ for haplotyper som støtter hendelsen. Disse kombineres til den betingede sannsynligheten $P(r|E)$ for en hendelseshypotese, E, som involverer en blanding av referanse- og kandidatsomatisk allel over en rekke mulige allelfrekvenser og multiplisert for å gi den betingede sannsynligheten $P(R|E)$ av å observere hele lesepileen. Derfra beregnes en TLOD-score som beviset på at en ALT-allel er tilstede i tumorprøven på et gitt locus.

DRAGEN for IDPE Dx bruker automatisk filtrering. Se [VCF-filmerknader for somatisk arbeidsflyt](#) på side 15 for mer informasjon.

Opprett en planlagt kjøring

Bruk følgende trinn for å sette opp en kjøring i Illumina Run Manager enten på NextSeq 550Dx eller ved hjelp av en nettleser på en nettverksdatamaskin. Bruk en nettleser på en nettverkstilkoblet datamaskin hvis du ønsker å importere prøvedata. Se Programvareveiledning for Illumina Run Manager for NextSeq 550Dx (dokumentnr. 200025239) for instruksjoner om tilgang Illumina Run Manager fra en nettverkstilkoblet datamaskin.

Det er to ulike måter å opprette en ny planlagt kjøring på:

- **Import Run** (Importer kjøring)—Bruk et prøveark fra en eksisterende kjøring som mal for en ny kjøring. Se Programvareveiledning for Illumina Run Manager for NextSeq 550Dx (dokumentnr. 200025239) for informasjon om hvordan du importerer en kjøring.
- **Creat Run** (opprett kjøring)—angi kjøringensparametere manuelt. Følgende instruksjoner beskriver hvordan du oppretter en kjøring.

MERK De nødvendige inndatafeltene i brukergrensesnittet er merket med en asteriks (*).

Applikasjon

1. Fra Runs (Kjøringer)-skjermens Planned (planlagt)-fane, velg **Create Run** (Opprett kjøring).
2. Velg DRAGEN for Illumina DNA Prep with Enrichment Dx-appen, og velg deretter **Next** (Neste).

Kjøringsinnstillinger

1. På skjermen Run Settings (Kjøreinstillinger), angi et unikt kjørenavn. Kjøringsnavnet identifiserer kjøringen fra sekvensering til analyse.
2. **[Valgfritt]** Angi en kjørebekrivelse for å identifisere kjøringen ytterligere.
3. Velg indeksadaptersett som brukes under bibliotekklargjøring.
4. Gå gjennom avleserlengden og endre om nødvendig. Read 1 (Avlesning 1) og Read 2 (Avlesning 2) har en standardverdi på 151 sykluser. Index 1 (Indeks 1) og Index 2 (Indeks 2) har en fast verdi på 10 sykluser og kan ikke endres.
5. **[Valgfritt]** Angi et bibliotek tube ID.
6. Velg **Next** (Neste).

Prøvedata

Prøvedata inkluderer prøve-ID, brønnposisjon (brønnposisjon for indeksplate) og biblioteknavn. Ved bruk av indeks A og B inkluderer brønnposisjon også plateidentifikator.

Det er to måter å angi prøvedata:

- **Import Samples** (Importer prøver)– Bruk en malfil som er tilgjengelig for nedlasting i skjermen Sample Data (Prøvedata).
- **Manually** (Manuell)– Angi prøvedataene rett i tabellen på skjermen Sample Data (Prøvedata).

Importere prøver

Når du planlegger en sekvenseringskjøring med en nettleser på en nettverkstilkoblet datamaskin, er en malfil (*.csv) tilgjengelig for nedlasting på skjermen Sample Data (Prøvedata). Malfilen er ikke tilgjengelig for nedlasting når du får tilgang til Illumina Run Manager gjennom operativsystemet NextSeq 550Dx. Gjør følgende for å angi prøvedata ved hjelp av funksjonen Import Samples (importer prøver).

MERK Fullfør Run Settings (Kjøreinnstillinger)-trinnene før du fortsetter.

1. Velg **Download Template** (Last ned mal) for å laste ned en tom CSV-fil.
2. Fra malfilen angir du prøvedata og lagrer filen. Biblioteknavn er valgfritt.

MERK Ved bruk av indeks A og B må dataene for kolonne B inkludere både plate- og brønnposisjon (brønnposisjon for indeksplate). Eksempel: A-A01, A-A02, A-A03.

3. Velg **Import Samples** (importerer prøver) og bla til malfilen som inneholder prøvedatainformasjonen fra forrige trinn.
4. Velg **Open** (åpne), **Proceed** (fortsett) og deretter **Next** (neste).

MERK Endring av prøve-ID før du velger Next (neste) kan føre til en feil. Fullfør kjøreoppsettet før du gjør endringer for å unngå feil.

Legge inn prøver manuelt

Bruk tabellen på skjermen Sample Data (Prøvedata) for å angi prøvedata manuelt.

1. Angi en unik prøve ID i feltet Sample ID (Prøve ID).
2. Bruk **Well Position** (brønnposisjon) (Index A eller Index B) eller **Plate - Well Position** (plate – brønnposisjon) (Index A&B) for å velge den tilknyttede indeksen for prøvene. Feltene i7 Index, Index 1, i5 Index og Index 2 fylles ut automatisk.
3. **[Valgfritt]** Skriv inn et biblioteknavn.
4. Legg til rader og gjenta trinn 1–3 etter behov til alle prøvene er lagt til i tabellen. Du kan legge til flere rader samtidig ved først å angi antall rader som skal legges til, og deretter velge +-ikonet. Du kan også fjerne rader ved å velge boksen ved siden av radnummeret, og deretter klikke på søppelikonet.
5. Velg **Next** (Neste).

Analyseinnstillinger

1. Velg ønsket analysearbeidsflyt:
 - FASTQ-generering
 - FASTQ- og VCF-generering for en kimlinjearbeidsflyt (manifestfil kreves)
 - FASTQ- og VCF-generering for en somatisk arbeidsflyt (manifestfil kreves)
2. **[Valgfritt] Generate ORA compressed FASTQs** (Generer ORA-komprimerte FASTQ-er) er aktivert som standard. FASTQ ORA-komprimering komprimerer FASTQ-filer opptil 5x uten tap sammenlignet med fastq.gz. Fjern merket for **Generate ORA compressed FASTQs** (Generer ORA-komprimerte FASTQ-er) hvis ukomprimerte data (fastq.gz) foretrekkes.
3. For kimlinje- og somatiske arbeidsflyter kreves en manifestfil. Bruk rullegardinmenyen **Manifest File Selection** (Manifestfilvalg) for å velge en manifestfil. Manifestet er en tabulordelt BED(*.bed)-fil som spesifiserer navnene og plasseringene til målrettede referanseregioner. For mer informasjon, se [Manifestfil på side 8](#).
4. **[Valgfritt]** For somatiske arbeidsflyter, bruk rullegardinmenyen **Noise File Selection** (Støyfilvalg) til å velge en systematisk støyfil.
En BED(*.bed.gz)-fil med stedsspesifikt støynivå kan spesifiseres for å filtrere ut systematisk støy. For mer informasjon, se [Støyfiltrering \(valgfritt\) på side 9](#).
5. Velg **Next** (Neste).

Kjøring Gjennomgå

1. På skjermen Review (Gjennomgå), se gjennom informasjonen for Run Settings (Kjøreinnstillinger), Sample Data (Prøvedata) og Analysis Settings (Analyseinnstillinger).
2. Velg **Save** (Lagre).
Kjøringen lagres i fanen Planned (Planlagt) på skjermen Runs (Kjøringer).

Innstillinger

Hvis du vil vise eller endre DRAGEN for IDPE Dx applikasjonsinnstillinger, velger du først Applications (Applikasjoner)-ikonet fra hovedskjermbildet. Velg deretter programmet du vil vise eller endre. En administratorkonto kreves for å endre innstillinger.

Konfigurasjon

Konfigurasjonsskjermen viser følgende programinnstillinger:

- **Library Prep Kits** (Biblioteksforberedelsessett)—Viser standard bibliotekforberedelsessett for appen. Denne innstillingen kan ikke endres.
- **Index adapter kits** (Indeksadaptersett)—Viser standard indeksadaptersett for appen. Denne innstillingen kan ikke endres.

- **Read lengths** (Leselengder)—Leselengder er satt til 151 for appen som standard, men kan endres under oppretting av kjøring.
- **Manifest and Noise Files** (Manifest og støyfiler)—Last opp og endre innstillinger for manifest- og støyfiler.
 - Velg **Upload File** (Last opp) fil for å laste opp filer for bruk i analyse.
 - Velg ikonet **Default** (Standard) for å angi filen som standardmanifest eller støyfil valgt under oppretting av kjøring når programmet er valgt.
 - Merk av for **Enabled** (Aktivert) for å angi at filen skal vises i rullegardinmenyen under kjøring.

Tillatelser

Bruk avmerkingsboksene på Permissions (Tillatelser)-skjermen til å administrere brukertilgang for appen.

Manifestfil

Ved bruk av DRAGEN for IDPE Dx, er en manifestfil nødvendig inndata for følgende arbeidsflyter:

- FASTQ og VCF-generering for en kimlinjearbeidsflyt
- FASTQ- og VCF-generering for en somatisk arbeidsflyt

Manifestet er en tabulator delt BED-tekstfil som bruker BED-formatet (*.bed) som spesifiserer navnene og plasseringene til målrettede referanseregioner. Hoveddelen av manifestfilen er avsnittet Regioner og skal inneholde følgende datakolonner:

Kolonne	Beskrivelse
Navn	Unikt brukerspesifisert navn for målet
Kromosom	Kromosomplassing (f.eks. chr10, chr5, osv.)
Start	1-basert indeks for målets startposisjon
Stopp	1-basert indeks for målets stopposisjon
Oppstrøms probelengde	Lengden på oppstrømsproben. For DRAGEN for IDPE Dx appen skal dette angis til 0.
Nedstrøms probelengde	Lengden på nedstrømsproben. For DRAGEN for IDPE Dx appen skal dette angis til 0.

MERK Et gyldig manifestfilformat kreves for analyse. DRAGEN vil stoppe analysen hvis manifestfilen er ugyldig.

Støyfiltrering (valgfritt)

Det systematiske støyfilteret er tilgjengelig for somatisk variantbetegnelse, og kan brukes til å redusere falske positive betegnelser ved å ta hensyn til stedsspesifikk støy. Den systematiske støyfilen genereres ved først å samle inn omtrent 50 normale prøver (fortrinnsvis spesifikke for panelet, bibliotekets forberedelse og sekvenser), og deretter deles summen av allefrekvenser under 30 % på hvert sted med tilstrekkelig dekning på det totale antallet prøver (Allelfrekvenser over 30 % antas å være kimlinjevarianter og ikke støy). Når støyverdiene er generert, filtreres somatiske varianter som detekteres på stedet.

Filteret kan brukes i Tumor-Normal-modus, men er spesielt nyttig for Tumor-Only-kjøringer der en matchet normal ikke er tilgjengelig. Den systematiske støyfilen må bruke en BED-fil som har filtypen (* .bed .gz) og må inneholde fire kolonner: Kromosom-, start-, slutt- og stedsspesifikke støynivåer for hver rad. Systematisk støyfiltrering er valgfritt.

Analyseutdataer

Kjøringer som pågår vises i kategorien Active (Aktiv). Fullførte kjøring vises i Completed (Fullført)-fanen. DRAGEN for IDPE Dx oppretter en analysemappe med unikt navn for hver analyse, som er separat fra mappen med sekvenseringsdata. Analysemappen inkluderer følgende informasjon:

- Manifestfil brukt
- Programvareversjon
- Prøve-ID
- Totalt justerte avlesninger
- Prosentandel av justerte avlesninger per prøve
- Antall SNV-er betegnet per prøve
- Antall indeler betegnet per prøve
- Dekningsstatistikk

Analyseutdatafiler

Analysemappens plassering spesifiseres av innstillingen External Storage for Analysis Results (Ekstern lagring for analyseresultater). Se Programvareveiledning for Illumina Run Manager for NextSeq 550Dx (dokumentnr. 200025239) for mer informasjon om innstillingen External Storage for Analysis Results (Ekstern lagring for analyseresultater).

På skjermen Run Details (kjøredetaljer) gir feltet External Location (Eksternplassering) banen for sekvenseringsdata. Analysemappens unike navn finnes i feltet Analysis Output Folder (Analyseutdatamappe) på skjermen Run Details (Kjøredetaljer). De nøyaktige filene som genereres avhenger av hvilken analysearbeidsflyt som brukes. Følgende analyseutdatafiler genereres av programmet.

MERK Hvis det oppstår en begrensingsfeil for maksimal filbanelengde ved tilgang til utdatafiler for analyse, prøv å flytte filen til en kortere baneplassering eller bruk en annen metode for å åpne filen.

Utdatafil	Beskrivelse
Sammendragsrapport for variant (*.pdf)	Inneholder et sammendrag av filinformasjon, programvareversjoner, prøveinformasjon, lesnivåstatistikk og SNV, innsetninger, slettinger og dekningsammendrag. Bare kimlinje og somatiske arbeidsflyter produserer en variantrapport.
FASTQ (*.fastq.gz eller *.fastq.ora)	Intermediære filer som inneholder kvalitetsscorede basebetegnelser. FASTQ-filer er primære inndata for innrettingstrinnet. Når ORA- komprimering er valgt, brukes filtypen *.fastq.ora.
Justering av BAM-filer (*.bam)	Inneholder justerte avlesninger for en gitt prøve.
Genom VCF-filer (*.gvcf.gz)	Inneholder genotypen for hver posisjon, enten betegnet som en variant eller betegnet som referanse.
VCF-filer (*.vcf.gz)	Inneholder varianter betegnet på hver posisjon.
Kjør metrikkrapport (*.csv)	Inneholder kvalitetsmålinger om kjøringen, inkludert ikke-indeksert totalutbytte og Q30-poengsum.

FASTQ-filer

FASTQ (*.fastq.gz, *.fastq.ora) er et tekstbasert filformat som inneholder basisbetegnelse og kvalitetsverdier per avlesning. Hver fil inneholder følgende informasjon:

- Den Prøveidentifikator
- Sekvensen
- Phred-kvalitetsscorene i et ASCII + 33-kodet format

Eksempelidentifikatoren formateres 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=AAAAAAAAA9#:<#<;<<<????#=#
```

BAM-filer

En BAM-fil (*.bam) er den komprimerte binære versjonen av en SAM-fil (sequence alignment map) som brukes til å representere justerte sekvenser på opptil 128 Mb. BAM-filer bruker filnavnformatet `SampleName_S#.bam`. # er prøvenummeret bestemt av rekkefølgen prøvene er oppført for kjøringen. I multinode-modus er S# satt til S1, uavhengig av rekkefølgen til prøven.

BAM-filer inneholder en overskriftsseksjon og en justeringsseksjon:

- **Header** (Overskrift)—Inneholder informasjon om hele filen, som eksempel navn, prøvelengde og justeringsmetode. Justeringer i justeringsdelen er knyttet til spesifikk informasjon i overskriftsdelen.
- **Alignments** (Justeringer)—Inneholder lesenavn, lesesekvens, lesekvalitet, justeringsinformasjon og egendefinerte tagger. Lesenavnet inkluderer kromosomet, startkoordinaten, justeringskvaliteten og samsvarsdeskriptorstrengen.

Justeringsdelen inneholder følgende informasjon for hvert avlesning eller avlesningspar:

- AS: Kvalitet for justering av paret ende.
- RG: Avlesningsgruppe, som angir antall avlesninger for en spesifikk prøve.
- BC: Strekkodetag, som indikerer den demultipleksede prøve-ID-en knyttet til avlesningen.
- SM: Kvalitet på justering med én ende.
- XC: Samsvarsbeskrivelsestreng.
- XN: Amplicon-navnemerke, som registrerer amplicon-ID-en knyttet til de avleste

BAM-indeksfiler (*.bam.bai) gir en indeks for den tilsvarende BAM-filen.

VCF-filer

Variant betegnelseformat (*.vcf) filer inneholder informasjon om varianter funnet på bestemte posisjoner i et referansegenom.

VCF-filoverskriften inkluderer VCF-filformatversjonen, variantbetegnelsesversjonen, og viser merknadene som brukes i resten av filen. VCF-overskriften inkluderer også referansegenomfilen og BAM-filen. Den siste linjen i overskriften inneholder kolonneoverskriftene for datalinjene. Hver av VCF-fildatalinjene inneholder informasjon om en enkelt variant.

Tabell 1 VCF-filoverskrifter

Overskrift	Beskrivelse
KROM	Kromosomet til referansegenomet. Kromosomer vises i samme rekkefølge som referanse-FASTA-filen.

Overskrift	Beskrivelse
POS	Enkeltbaseposisjonen til varianten i referansekromosomet. For enkelt nukleotidvarianter (SNV) er denne posisjonen referansebasen med varianten. For indels er denne posisjonen referansegrunnet umiddelbart før varianten.
ID	rs (referanse SNP) nummer for SNP hentet fra <code>dbSNP.txt</code> , hvis aktuelt. Hvis det finnes flere rs-nummer på dette stedet, er listen avgrenset med semikolon. Hvis en dbSNP-oppføring ikke eksisterer på denne posisjonen, brukes en manglende verdemarkør ('.').
REF	Referansegenotypen. For eksempel er en sletting av en enkelt T representert som referanse TT og alternativ T. En enkel nukleotidvariant A til T er representert som referanse A og alternativ T.
ALT	Allelene som skiller seg fra referansen leses. For eksempel er en innsetting av en enkelt T representert som referanse A og alternativ AT. En enkelt nukleotidvariant A til T er representert som referanse A og alternativ T.
KVAL	En Phred-skalert kvalitetspoeng tildelt av varianten som betegner. Høyere scoringer indikerer høyere tillit til varianten og lavere sannsynlighet for feil. For en kvalitetsscore på Q er den estimerte sannsynligheten for en feil $10^{-(Q/10)}$. For eksempel har settet med Q30-betegnelser en feilfrekvens på 0,1 %. Mange variantanropere tildeler kvalitetspoeng basert på deres statistiske modeller, som er høye i forhold til observert feilprosent.

Tabell 2 VCF-filmerknader for kimlinjearbeidsflyt

Overskrift	Beskrivelse
FILTER	<p>Hvis alle filtre er bestått, skrives PASS (bestått) i filterkolonnen. Mulige FILTER-oppføringer inkluderer:</p> <ul style="list-style-type: none"> • DRAGENSnpHardQUAL—Brukes hvis SNP-variant QUAL-poengsum ikke når terskelen • DRAGENIndelHardQUAL—Brukes hvis indel variant QUAL-poengsum ikke når terskelen • LowDepth—stedet filtreres fordi dekningsdybden ikke når terskelen • LowGQ—stedet filtrert fordi genotypkvaliteten ikke oppfyller terskelen • PloidyConflict—Genotypebetegnelse fra variantbetegnelser som ikke stemmer overens med kromosomploidi • base_quality—stedet ble filtrert fordi median basekvalitet for alt-avlesninger på dette lokuset ikke når terskelen • filtered_reads—stedet er filtrert fordi en for stor del av avlesningene har blitt filtrert ut • fragment_length—stedet filtrert fordi den absolutte forskjellen mellom median fragmentlengden for alt-avlesninger og median-fragmentlengden til ref-avlesningene på dette stedet overskrider terskelen • low_depth—stedet ble filtrert fordi avlesningsdybden er for lav • low_frac_info_reads—stedet filtrert fordi andelen av informative avlesninger er under terskelen • low_normal_depth—stedet filtrert fordi den normale prøvelesedybden er for lav • long_indel—Stedet ble filtrert fordi indellengden er for lang • mapping_quality—Stedet filtrert fordi median kartleggingskvalitet for alt-avlesninger på dette lokuset ikke når terskelen • multiallelic—Sted filtrert fordi mer enn to alt-alleler passerer tumor LOD • non_homref_normal—stedet filtrert fordi den normale prøvegenotypen ikke er homozygot referanse • no_reliable_supporting_read—Stedet filtrert fordi det ikke finnes noen pålitelig støtte for somatisk avlesning • panel_of_normals—Sett i minst ett utvalg i panelet av normaler vcf • read_position—Stedet filtrert fordi medianen av avstander mellom start/slutt av avlesning og dette lokuset er under terskelen • RMxNRepeatRegion—Stedet filtrert fordi hele eller deler av variantallelen er en repetisjon av referansen • strand_artifact—Stedet filtrert på grunn av alvorlig strengsavvik • str_contraction—Stedet filtrert på grunn av mistenkt PCR-feil der alt-allelen er én gjentatt enhet mindre enn referansen • too_few_supporting_reads—Stedet filtrert fordi det er for få støttende avlesninger i tumorprøven • weak_evidence—Somatisk variantscore når ikke terskelen

Overskrift	Beskrivelse
INFO	<p>Mulige INFO-oppføringer inkluderer:</p> <ul style="list-style-type: none"> • AC—Alleltelling i genotyper for hver ALT-allel, i samme rekkefølge som oppført. • AF—Allelfrekvens for hver ALT-allel, i samme rekkefølge som oppført. • AN—Det totale antallet alleler i betegnede genotyper. • DB—dbSNP-medlemskap. • FS—Phred-skalert p-verdi ved å bruke Fisher's eksakte test for å oppdage strengskjevhet. • QD—Variant tillit/kvalitet etter dybde. • R2_5P_bias—Poeng basert på kameratavvik og avstand fra 5 prime ende. • SOR—symmetrisk oddsforhold på 2x2 beredskapstabell for å oppdage strengskjevhet. • DP—Omtrentlig avlesningsdybde (informativ og ikke-informativ); noen avlesninger kan ha blitt filtrert basert på mapq osv. • END—Stoppposisjon for intervallet. • FractionInformativeReads—Brøkdelen av informative avlesninger av det totale antallet avlesninger. • MQ—RMS-kartkvalitet. • MQRankSum—Z-score fra Wilcoxon rangsumtest av Alt vs. Ref avlesningskartleggingskvaliteter. • ReadPosRankSum—Z-score fra Wilcoxon rangsumtest av Alt vs. Ref avlesningsposisjonsavvik. • SOMATISK—Minst én variant i denne stillingen er somatisk.
FORMAT	<p>Formatkolonnen viser felt atskilt med kolon. For eksempel GT:GQ. Tilgjengelige felt inkluderer:</p> <ul style="list-style-type: none"> • AD—Alleliske dybder (teller bare informative avlesninger av de totale avlesningene) for ref- og alt-allelene i den oppførte rekkefølgen. • AF—Allelbrøker for alt-allelene i den oppførte rekkefølgen. • DP—Omtrentlig avlesningsdybde (avlesninger med MQ=255 eller med dårlige kamerater filtreres). • F1R2—Antall avlesninger i F1R2-parorientering som støtter hvert allel. • F2R1—Antall avlesninger i F2R1-parorientering som støtter hvert allel. • GT—Genotype. 0 tilsvarer referansebasen, 1 tilsvarer den første oppføringen i ALT-kolonnen, og så videre. Skråstreken (/) indikerer at ingen faseinformasjon er tilgjengelig. • MB—Per-prøve komponentstatistikk for å oppdage partner-bias. • PS—Fysisk fase-ID-informasjon, der hver unik ID i en gitt prøve (men ikke på tvers av prøver) kobler sammen poster innenfor en fasegruppe. • SB—Per-prøve komponentstatistikk som omfatter Fisher's Exact Test for å oppdage strengskjevhet. • SQ—Somatisk kvalitet.

Overskrift	Beskrivelse
SAMPLE (PRØVE)	Eksempelkolonnen gir verdiene angitt i FORMAT-kolonnen.

Tabell 3 VCF-filmerknader for somatisk arbeidsflyt

Overskrift	Beskrivelse
FILTER	<p>Hvis alle filtre er bestått, skrives PASS (bestått) i filterkolonnen. Mulige FILTER-oppføringer inkluderer:</p> <ul style="list-style-type: none"> • base_quality—stedet ble filtrert fordi median basekvalitet for alt-avlesninger på dette lokuset ikke når terskelen • filtered_reads—Stedet er filtrert fordi en for stor del av lesningene er filtrert ut • fragment_length—stedet filtrert fordi den absolutte forskjellen mellom median fragmentlengden for alt-avlesninger og median-fragmentlengden til ref-avlesningene på dette stedet overskrider terskelen • low_depth—stedet ble filtrert fordi avlesningsdybden er for lav • low_frac_info_reads—stedet filtrert fordi andelen av informative avlesninger er under terskelen • low_normal_depth—stedet filtrert fordi den normale prøvelesedybden er for lav • long_indel—Stedet ble filtrert fordi indellengden er for lang • mapping_quality—Stedet filtrert fordi median kartleggingskvalitet for alt-avlesninger på dette lokuset ikke når terskelen • multiallelic—Sted filtrert fordi mer enn to alt-alleler passerer tumor LOD • non_homref_normal—stedet filtrert fordi den normale prøvegenotypen ikke er homozygot referanse • no_reliable_supporting_read—Stedet filtrert fordi det ikke finnes noen pålitelig støtte for somatisk avlesning • panel_of_normals—Sett i minst ett utvalg i panelet av normaler vcf • read_position—Stedet filtrert fordi medianen av avstander mellom start/slutt av avlesning og dette lokuset er under terskelen • RMxNRepeatRegion—Stedet filtrert fordi hele eller deler av variantallelen er en repetisjon av referansen • strand_artifact—Stedet filtrert på grunn av alvorlig strengsavvik • str_contraction—Stedet filtrert på grunn av mistenkt PCR-feil der alt-allelen er én gjentatt enhet mindre enn referansen • too_few_supporting_reads—Stedet filtrert fordi det er for få støttende avlesninger i tumorprøven • weak_evidence—Somatisk variantscore når ikke terskelen • systematic_noise—Sted filtrert basert på bevis på systematisk støy i normaler

Overskrift	Beskrivelse
INFO	<p>Mulige INFO-oppføringer inkluderer:</p> <ul style="list-style-type: none"> • DP—Omtrentlig avlesningsdybde (informativ og ikke-informativ); noen avlesninger kan ha blitt filtrert basert på mapq osv. • END—Stoppposisjon for intervallet. • FractionInformativeReads—Brøkdelen av informative avlesninger av det totale antallet avlesninger. • MQ—RMS-kartkvalitet. • MQRankSum—Z-score fra Wilcoxon rangsumtest av Alt vs. Ref avlesningskartleggingskvaliteter. • ReadPosRankSum—Z-score fra Wilcoxon rangsumtest av Alt vs. Ref avlesningsposisjonsavvik. • AQ—Systematisk støyscore. • hotspot—kjent somatisk nettsted, brukt for å øke tilliten til samtalen. • SOMATISK—Minst én variant i denne stillingen er somatisk.
FORMAT	<p>Formatkolonnen viser felt atskilt med kolon. For eksempel GT:GQ. Tilgjengelige felt inkluderer:</p> <ul style="list-style-type: none"> • AD—Alleliske dybder (teller bare informative avlesninger av de totale avlesningene) for ref- og alt-allelene i den oppførte rekkefølgen. • AF—Allelbrøker for alt-alleler i den oppførte rekkefølgen. • DP—Omtrentlig avlesningsdybde (avlesninger med MQ=255 eller med dårlige kamerater filtreres). • F1R2—Antall avlesninger i F1R2-parorientering som støtter hvert allel. • F2R1—Antall avlesninger i F2R1-parorientering som støtter hvert allel. • GP—Phred-skalerte posteriore sannsynligheter for genotyper som definert i VCF-spesifikasjonen. • GQ—Genotype kvalitet. • GT—Genotype. 0 tilsvarer referansebasen, 1 tilsvarer den første oppføringen i ALT-kolonnen, og så videre. Skråstreken (/) indikerer at ingen faseinformasjon er tilgjengelig. • MB—Per-prøve komponentstatistikk for å oppdage partner-bias. • PL—Normaliserte, Phred-skalerte sannsynligheter for genotyper som definert i VCF-spesifikasjonen. • PRI—Phred-skalerte tidligere sannsynligheter for genotyper. • PS—Fysisk fase-ID-informasjon, der hver unik ID i en gitt prøve (men ikke på tvers av prøver) kobler sammen poster innenfor en fasegruppe. • SB—Per-prøve komponentstatistikk som omfatter Fisher's Exact Test for å oppdage strengskjevhet. • SQ—Somatisk kvalitet.
SAMPLE (PRØVE)	Eksempelkolonnen gir verdiene angitt i FORMAT-kolonnen.

Genom VCF-filer

Genom VCF (*.gvcf.gz)-filer følger et sett med konvensjoner for å representere alle steder i genomet i et rimelig kompakt format. gVCF-filene inkluderer alle steder innenfor området av interesse i en enkelt fil for hver prøve. gVCF-filen viser ingen betegnelse på posisjoner som ikke passerer alle filtre. En genotype (GT)-kode av ./ indikerer ingen betegnelse.

Requeue Analysis (Sett analyse tilbake i kø)

Du kan sette en analyse tilbake i kø hvis den har blitt stoppet, hvis analysen mislyktes, eller hvis du ønsker å analysere en kjøring på nytt med andre innstillinger. Gjør følgende for å sette analyse tilbake i kø:

1. Fra Run (Kjøring)-skjermen, velg Completed (Fullført)-fanen og deretter kjøringens navn som skal analyseres på nytt.
Hvis Requeue Analysis (Sett analyse tilbake i kø) tidligere ble utført, velger du kjøringens navn for den overordnede kjøringen.
2. Fra Run Details (Kjøringdetaljer)-skjermen, etter Sequencing Information (sekvenseringsinformasjon), velg **Requeue Analysis** (Sett analyse tilbake i kø).
3. Velg et alternativ:
 - Sett analyse tilbake i kø uten endringer
 - Rediger kjøringens innstillinger og sett analyse tilbake i kø
 - Sett analysen tilbake i kø med et annet program
4. Bekreft at plasseringen der sekvenseringsdataene for øyeblikket befinner seg, er angitt i feltet **Sequencing data file path** (Filbane for sekvenseringsdata).

MERK Banen til sekvenseringsdataene skal samsvare med banen i External Storage for Analysis Results (Ekstern lagring for analyseresultater)-innstillingen. Se Programvareveiledning for Illumina Run Manager for NextSeq 550Dx (dokumentnr. 200025239) for informasjon om endring av den eksterne lagringsbanen.

5. Angi årsak for ny analyse.
6. Velg **Requeue Analysis** (Sett analyse tilbake i kø).
7. Rediger ønskede endringer i innstillingene for Run Settings (Kjøreinnstillinger), Sample Data (Prøvedata) og Analysis (Analyse).
8. Velg **Save** (Lagre). Analysen starter med gjeldende analyseparametre.

Teknisk hjelp

Kontakt teknisk støtte hos Illumina for teknisk hjelp.

Nettsted: www.illumina.com

E-post: techsupport@illumina.com

Sikkerhetsdatablad (SDS)—Tilgjengelige på Illuminas nettsted på support.illumina.com/sds.html.

Produktdokumentasjon—Tilgjengelig for nedlasting fra support.illumina.com.



Illumina
5200 Illumina Way
San Diego, California, 92122 USA
+1 800 809 ILMN (4566)
+1 858 202 4566 (utenfor Nord-Amerika)
techsupport@illumina.com
www.illumina.com



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

TIL IN VITRO-DIAGNOSTISK BRUK.

© 2023 Illumina, Inc. Alle rettigheter forbeholdt.

illumina[®]