

# DRAGEN for Illumina DNA Prep med Enrichment Dx

Produktdokumentasjon for NovaSeq 6000Dx

ILLUMINA-PROPRIETÆR

Dokumentnr. 200014776 v02

September 2022

TIL IN VITRO-DIAGNOSTISK BRUK

## Revisjonshistorikk

Dokument	Dato	Beskrivelse av endring
200014776 v02	September 2022	Korrigert format for manifestfil fra tekst (*.txt) til BED (*.bed) i instruksjoner for opprette løp Korrigerte konsensus VCF-filer til VCF-filer i analyseutdataseksjonen.
200014776 v01	August 2022	Lagt til: Innstillinger delen. Systematisk støyfiltreringsseksjon. Oppdaterte instruksjoner for oppretting av kjøring for å inkludere flere detaljer. Rettet skrivefeil og grammatiske feil. Spesifisert at instruksjoner er ment for appen når den brukes med NovaSeq 6000Dx-instrumentet. Oppdatert informasjon om innholdet i VCF-utdatafilen.
200014776 v00	Mars 2022	Første versjon.

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).

© 2022 Illumina, Inc. Alle rettigheter forbeholdt.

Alle varemerker tilhører Illumina, Inc. eller deres respektive eiere. Ytterligere informasjon om varemerker finner du på [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).

# Innholdsfortegnelse

Revisjonshistorikk .....	ii
Oversikt .....	1
Analysemetoder .....	1
Opprett en kjøring .....	4
Innstillinger .....	5
Analyseutdataer .....	7
FASTQ-filer .....	8
BAM filer .....	8
VCF-filer .....	9
Vise analyseresultater .....	14
<b>Teknisk hjelp .....</b>	<b>15</b>

# Oversikt

Den DRAGEN™ for Illumina® DNA Prep with Enrichment Dx Applikasjonen utfører demultipleksing, FASTQ generasjon, avlesningskartlegging og justering til et referansegenom og variantbetegnelse avhengig av valgt analysearbeidsflyt.

## Analysemetoder

DRAGEN for Illumina DNA Prep med Enrichment Dx utfører demultipleksing, FASTQ-generering, lesekartlegging og justering til et referansegenom avhengig av den valgte arbeidsflyten:

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

## 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. Når du bruker ORA (\*.ora)-formatet, bevares md5-sjekksummen av FASTQ-innholdet etter en komprimerings- og dekompresjonssyklus for å sikre en tapsfri komprimering.

## DNA-kartlegging og justering

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.

De spesifikke verdiene som er valgt for scoringer i denne algoritmen indikerer hvordan man balanserer, for en justering med flere mulige tolkninger, muligheten for en indel i motsetning til en eller flere SNP-er, eller preferansen for en justering uten klipping. Standard DRAGEN-scoringsverdier er rimelige for å justere moderat lengdeavlesninger til et helt menneskelig referansegenom for variantbetegnelsesapplikasjoner. Ethvert sett med Smith-Waterman-scoringsparametere representerer en upresis modell av genomisk mutasjon og sekvenseringsfeil. Ulikt innstilte justeringsscoringsverdier kan være mer passende for enkelte applikasjoner.

## DRAGEN Kimlinje variantbetegnelse

DRAGEN Kimlinje liten variantbetegnelser tar kartlagte og justerte DNA-avlesninger som input og betegner SNP-er og indeler gjennom en kombinasjon av kolonnevis deteksjon og lokal *de novo*-sammenstilling av haplotyper.

Betegnelige referanseregioner identifiseres først med tilstrekkelig innrettingsdekning. Innenfor disse referanseområdene identifiserer en rask skanning av de sorterte avlesningene aktive regioner, som er sentrert rundt 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.

I gVCF-modusen som brukes for skalerbar multisample-variantbetegnelse, kan DRAGEN kimlinje liten variantbetegnelse kjøres per prøve for å generere en mellomliggende genomisk variant-anropsfil (gVCF). gVCF kan deretter brukes til effektiv felles genotyping av flere prøver, noe som muliggjør rask inkrementell behandling av prøver og skalering til store kohortstørrelser.

Fordi DRAGEN Kimlinje liten variantbetegnelse har algoritmer som gjør den i stand til effektivt å skille korrelerte feil fra sanne varianter, filtreringsreglene er veldig enkle.

## DRAGEN Somatisk variantbetegnelse

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.

Betegnelige referanseregioner identifiseres først med tilstrekkelig innrettingsdekning. Innenfor disse referanseregionene identifiserer en skanning av de sorterte avlesningene aktive regioner, som er sentrert rundt pileup-søyler med bevis på en variant i svulsten. 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 TLOD poengsummen, DRAGEN Somatic Small Variant Ringer først etter referanseposisjon for hver kandidat somatisk hendelse, samt referansehendelsen over det aktive området. Den betingede sannsynligheten  $P(r|e)$  for å observere hver overlappende lesning er estimert som den maksimale  $P(r|H)$  for haplotyper som støtter arrangementet. 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 lese-pilen. Derfra beregnes en TLOD-score som beviset på at en ALT-allel er tilstede i tumorprøven på et gitt locus.

# Opprett en kjøring

Bruk følgende trinn for å sette opp en kjøring i Illumina Run Manager enten på NovaSeq 6000Dx eller ved hjelp av en nettleser på en nettverksdatamaskin. Prøvedata kan legges inn manuelt eller ved å importere et prøveark.

## Innstillinger for applikasjon og kjøring

1. Fra Runs (Kjøringer) skjermen, velger du **Create Run** (Opprett kjøring).
2. Velg DRAGEN for Illumina DNA Prep med Enrichment Dx-appen, og velg deretter **Next** (Neste).
3. På skjermbildet Run Settings (Kjøreinstillinger) skriver du inn et kjørningsnavn. Kjørningsnavnet identifiserer kjøringen fra sekvensering til analyse.
4. **[Valgfritt]** Angi en kjørebekrivelse for å identifisere kjøringen ytterligere.
5. Forsikre deg om at bibliotekforberedelsessettet som er valgt, er et Illumina DNA Prep with Enrichment Dx bibliotekforberedelsessett.
6. Velg ønsket indeksadaptersett.
7. Skriv inn Read Length (Avlesningslengden).  
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.
8. **[Valgfritt]** Angi et bibliotek tube ID.
9. Velg **Next** (Neste).

## Prøvedata

Bruk tabellen på skjermen Sample Data (Prøver Data) for å legge inn prøveinformasjon manuelt. Alternativt, Velg **Import Samples** (Importer prøver) prøver for å laste opp eksempelinformasjon. For informasjon om import av eksempelinformasjon, se [Importere prøver på side 5](#).

Legge inn prøver manuelt

1. Angi en unik prøve ID i feltet Sample ID (Prøve ID).
2. Bruk **Plate - Well Position** (Plate - brønnposisjon) for å velge brønnposisjon.  
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.
5. Velg **Next** (Neste).



## Importere prøver

En mal (\*.csv) er tilgjengelig for nedlasting på skjermen Sample Data (Prøver Data) når du planlegger en kjøring i Illumina Run Manager ved hjelp av en nettleser på en nettverksbasert datamaskin.

1. Velg **Download Template** (Last ned mal) for å laste ned en tom CSV-fil.
2. Skriv inn eksempelinformasjonen fra CSV-filen og lagre filen.  
Eksempelarket CSV-filen inneholder følgende datakolonner: Sample ID (Prøve-ID), Plate - Well Position (Plate - brønnposisjon), **valgfritt** Library Name (Biblioteknavn).
3. Velg **Import Samples** (Importer prøver) og bla til plasseringen av CSV-filen.
4. Velg **Next** (Neste).

## Analyseinnstillinger

1. Velg ønsket analysearbeidsflyt:
  - FASTQ-generering
  - Germline FASTQ og VCF generasjon for en kimlinjearbeidsflyt
  - Somatisk FASTQ- og VCF-generering for en somatisk arbeidsflyt
2. **[Valgfritt]** Hvis ønskelig, merk av for **Generate ORA compressed FASTQs** (Generer ORA-komprimerte FASTQ-er) for å aktivere FASTQ ORA komprimering.
3. **[Arbeidsflyter for VCF-generering]** Bruk ikonet **Manifest File Selection** (Manifest Filvalg) rullegardinmenyen for å velge en manifestfil.  
Det kreves en manifestfil for DRAGEN for Illumina DNA Prep med Enrichment Dx. Manifestet er en tabulator delt BED fil (\*.bed) som spesifiserer navnene og plasseringene til målrettede referanseregioner.
4. **[Arbeidsflyt for somatisk FASTQ- og VCF-generering]** Bruk ikonet **Noise File Selection** (Støyfilvalg) rullegardin for å velge en støyfil.  
En BED-fil med stedsspesifikt støynivå kan spesifiseres for å filtrere ut systematisk støy. For mer informasjon, se [Støyfiltrering på side 6](#).
5. Velg **Next** (Neste).

## Kjøring Gjennomgå

1. På skjermen Review (Gjennomgå), se gjennom informasjonen som er angitt på skjermene 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

Velg appen på skjermen Applications (Applikasjons) for å se gjeldende innstillinger og 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 avmerkbingsboksene på Permissions (Tillatelser)-skjermen til å administrere brukertilgang for appen.

## Støyfiltrering

Systematisk støyfiltrering er tilgjengelig når du bruker den somatiske arbeidsflyten. Filteret kan brukes i Tumor-Normal-modus, men er spesielt nyttig for Tumor-Only-kjøringer der en matchet normal ikke er tilgjengelig.

Den systematiske støy BED bør genereres fra vanlige prøver. Det anbefales å bygge systematiske støyfiler som er spesifikke for biblioteksforberedelser, sekvenseringssystem og panel. Det anbefales å bruke omtrent 50 normale prøver for generering av støyfiler.

# Analyseutdataer

DRAGEN for Illumina DNA Prep med Enrichment Dx lagrer følgende informasjon i analysemappen. Bare kimlinje og somatiske arbeidsflyter produserer en PDF.

- 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

Følgende utdatafiler genereres av applikasjonen. De nøyaktige filene som genereres avhenger av hvilken analysearbeidsflyt som brukes. Utdatafiler ligger i analysemappen.

Utdatafil	Beskrivelse
FASTQ (*.fastq.gz eller *.fastq.ora)	Intermediære filer som inneholder kvalitetsscorede basebetegnelser. FASTQ-filer er primære inndata for innrettingstrinnet. Hvis ORA-komprimering er valgt, gjenspeiler filnavnet dette.
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 kalt på hver posisjon.
Kjør metrikkrapport (*.csv)	Inneholder kvalitetsmålinger om kjøringen, inkludert 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
- Et plusstegn (+)
- 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=AAAAAAAAAAA9#:<#<;<<<?????#=#
```

## 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:

- Overskrift—Inneholder informasjon om hele filen, som eksempel navn, prøvelengde og justeringsmetode. Justeringer i justeringsdelen er knyttet til spesifikk informasjon i overskriftsdelen.
- 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 leste BAM indeksfilene (\*.bam.bai) gir en indeks over 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 og varianten som betegner, 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.
POS	Enkeltnukleotidposisjonen til varianten i referansegenomet. For enkeltnukleotidvarianter (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 verdimerker ('.').
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 enkeltnukleotidvariant 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-filanmerkninger

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

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

Overskrift	Beskrivelse
INFO	<p>Mulige INFO-oppføringer for kimlinjearbeidsflyt inkluderer:</p> <ul style="list-style-type: none"> <li>• <b>AC</b>—Alleltelling i genotyper for hver ALT-allel, i samme rekkefølge som oppført.</li> <li>• <b>AF</b>—Allelfrekvens for hver ALT-allel, i samme rekkefølge som oppført.</li> <li>• <b>AN</b>—Det totale antallet alleler i betegnede genotyper.</li> <li>• <b>DB</b>—dbSNP-medlemskap.</li> <li>• <b>FS</b>—Phred-skalert p-verdi ved å bruke Fishers eksakte test for å oppdage strengskjevhet.</li> <li>• <b>QD</b>—Variant tillit/kvalitet etter dybde.</li> <li>• <b>R2_5P_bias</b>—Poeng basert på kameratavvik og avstand fra 5 prime ende.</li> <li>• <b>SOR</b>—symmetrisk oddsforhold på 2x2 beredskapstabell for å oppdage strengskjevhet.</li> <li>• <b>DP</b>—Omtrentlig avlesningsdybde (informativ og ikke-informativ); noen avlesninger kan ha blitt filtrert basert på mapq osv.</li> <li>• <b>END</b>—Stoppposisjon for intervallet.</li> <li>• <b>FractionInformativeReads</b>—Brøkdelen av informative avlesninger av det totale antallet avlesninger.</li> <li>• <b>MQ</b>—RMS-kartkvalitet.</li> <li>• <b>MQRankSum</b>—Z-score fra Wilcoxon rangsumtest av Alt vs. Ref avlesningskartleggingskvaliteter.</li> <li>• <b>ReadPosRankSum</b>—Z-score fra Wilcoxon rangsumtest av Alt vs. Ref avlesningsposisjonsavvik.</li> <li>• <b>SOMATISK</b> – Minst én variant i denne stillingen er somatisk.</li> </ul> <p>Somatisk arbeidsflyt mulige INFO-oppføringer inkluderer:</p> <ul style="list-style-type: none"> <li>• <b>DP</b>—Omtrentlig avlesningsdybde (informativ og ikke-informativ); noen avlesninger kan ha blitt filtrert basert på mapq osv.</li> <li>• <b>END</b>—Stoppposisjon for intervallet.</li> <li>• <b>FractionInformativeReads</b>—Brøkdelen av informative avlesninger av det totale antallet avlesninger.</li> <li>• <b>MQ</b>—RMS-kartkvalitet.</li> <li>• <b>MQRankSum</b>—Z-score fra Wilcoxon rangsumtest av Alt vs. Ref avlesningskartleggingskvaliteter.</li> <li>• <b>ReadPosRankSum</b>—Z-score fra Wilcoxon rangsumtest av Alt vs. Ref avlesningsposisjonsavvik.</li> <li>• <b>AQ</b>—Systematisk støyscore.</li> <li>• <b>hotspot</b>—kjent somatisk nettsted, brukt for å øke tilliten til samtalen.</li> <li>• <b>SOMATISK</b> – Minst én variant i denne stillingen er somatisk.</li> </ul>



Overskrift	Beskrivelse
FORMAT	<p>Formatkolonnen viser felt atskilt med kolon. For eksempel GT:GQ.</p> <p>Tilgjengelige felt for kimlinjearbeidsflyt inkluderer:</p> <ul style="list-style-type: none"> <li>• <b>AD</b>—Alleliske dybder (teller bare informative avlesninger av de totale avlesningene) for ref- og alt-allelene i den oppførte rekkefølgen.</li> <li>• <b>AF</b>—Allelbrøker for alt-alleler i den oppførte rekkefølgen.</li> <li>• <b>DP</b>—Omtrentlig avlesningsdybde (avlesninger med MQ=255 eller med dårlige kamerater filtreres).</li> <li>• <b>F1R2</b>—Antall avlesninger i F1R2-parorientering som støtter hvert allel.</li> <li>• <b>F2R1</b>—Antall avlesninger i F2R1-parorientering som støtter hvert allel.</li> <li>• <b>GP</b>—Phred-skalerte posteriore sannsynligheter for genotyper som definert i VCF-spesifikasjonen.</li> <li>• <b>GQ</b> — Genotype kvalitet.</li> <li>• <b>GT</b>—Genotype. 0 tilsvarer referansebasen, 1 tilsvarer den første oppføringen i ALT-kolonnen, og så videre. Skråstreken (/) indikerer at ingen faseinformasjon er tilgjengelig.</li> <li>• <b>MB</b>—Per-prøve komponentstatistikk for å oppdage partner-bias.</li> <li>• <b>PL</b>—Normaliserte, Phred-skalerte sannsynligheter for genotyper som definert i VCF-spesifikasjonen.</li> <li>• <b>PRI</b>—Phred-skalerte tidligere sannsynligheter for genotyper.</li> <li>• <b>PS</b>—Fysisk fase-ID-informasjon, der hver unik ID i en gitt prøve (men ikke på tvers av prøver) kobler sammen poster innenfor en fasegruppe.</li> <li>• <b>SB</b>—Per-prøve komponentstatistikk som omfatter Fisher's Exact Test for å oppdage strengskjevhet.</li> <li>• <b>SQ</b>—Somatisk kvalitet.</li> </ul> <p>Tilgjengelige felt for somatisk arbeidsflyt inkluderer:</p> <ul style="list-style-type: none"> <li>• <b>AD</b>—Alleliske dybder (teller bare informative avlesninger av de totale avlesningene) for ref- og alt-allelene i den oppførte rekkefølgen.</li> <li>• <b>AF</b>—Allelbrøker for alt-alleler i den oppførte rekkefølgen.</li> <li>• <b>DP</b>—Omtrentlig avlesningsdybde (avlesninger med MQ=255 eller med dårlige kamerater filtreres).</li> <li>• <b>F1R2</b>—Antall avlesninger i F1R2-parorientering som støtter hvert allel.</li> <li>• <b>F2R1</b>—Antall avlesninger i F2R1-parorientering som støtter hvert allel.</li> <li>• <b>GT</b>—Genotype. 0 tilsvarer referansebasen, 1 tilsvarer den første oppføringen i ALT-kolonnen, og så videre. Skråstreken (/) indikerer at ingen faseinformasjon er tilgjengelig.</li> <li>• <b>MB</b>—Per-prøve komponentstatistikk for å oppdage partner-bias.</li> <li>• <b>PS</b>—Fysisk fase-ID-informasjon, der hver unik ID i en gitt prøve (men ikke på tvers av prøver) kobler sammen poster innenfor en fasegruppe.</li> <li>• <b>SB</b>—Per-prøve komponentstatistikk som omfatter Fisher's Exact Test for å oppdage strengskjevhet.</li> <li>• <b>SQ</b>—Somatisk kvalitet.</li> </ul>

Overskrift	Beskrivelse
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.

## Vise analyseresultater

Kjøringer som pågår vises i kategorien Active (Aktiv). Fullførte kjøring vises i Completed (Fullført)-fanen. Se [NovaSeq 6000Dx produktdokumentasjon \(dokumentnr. 200010105\)](#) for mer informasjon om visning av resultater.

# Teknisk hjelp

Kontakt teknisk støtte hos Illumina for teknisk hjelp.

**Nettsted:** [www.illumina.com](http://www.illumina.com)  
**E-post:** [techsupport@illumina.com](mailto:techsupport@illumina.com)

## Telefonnumre til Illumina teknisk støtte

Region	Gratis	Internasjonalt
Australia	+61 1800 775 688	
Østerrike	+43 800 006249	+43 1 9286540
Belgia	+32 800 77 160	+32 3 400 29 73
Canada	+1 800 809 4566	
Kina		+86 400 066 5835
Danmark	+45 80 82 01 83	+45 89 87 11 56
Finland	+358 800 918 363	+358 9 7479 0110
Frankrike	+33 8 05 10 21 93	+33 1 70 77 04 46
Tyskland	+49 800 101 4940	+49 89 3803 5677
Hongkong, Kina	+852 800 960 230	
India	+91 8006500375	
Indonesia		0078036510048
Irland	+353 1800 936608	+353 1 695 0506
Italia	+39 800 985513	+39 236003759
Japan	+81 0800 111 5011	
Malaysia	+60 1800 80 6789	
Nederland	+31 800 022 2493	+31 20 713 2960
New Zealand	+64 800 451 650	
Norge	+47 800 16 836	+47 21 93 96 93
Filippinene	+63 180016510798	
Singapore	1 800 5792 745	
Sør-Korea	+82 80 234 5300	

Region	Gratis	Internasjonalt
Spania	+34 800 300 143	+34 911 899 417
Sverige	+46 2 00883979	+46 8 50619671
Sveits	+41 800 200 442	+41 56 580 00 00
Taiwan, Kina	+886 8 06651752	
Thailand	+66 1800 011 304	
Storbritannia	+44 800 012 6019	+44 20 7305 7197
USA	+1 800 809 4566	+1 858 202 4566
Vietnam	+84 1206 5263	

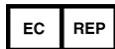
**Sikkerhetsdatablad** – Tilgjengelige på Illuminas nettsted på [support.illumina.com/sds.html](https://support.illumina.com/sds.html).

**Produktdokumentasjon** – Tilgjengelig for nedlasting fra [support.illumina.com](https://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

CE



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

**Australsk sponsor**

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

TIL IN VITRO-DIAGNOSTISK BRUK

© 2022 Illumina, Inc. Alle rettigheter forbeholdt.

**illumina**<sup>®</sup>