

# DRAGEN for Illumina DNA Prep with Enrichment Dx

Produktdokumentation for NovaSeq 6000Dx

## Revisionshistorik

Dokument	Dato	Beskrivelse af ændring
200014776 v02	September 2022	Rettet formatet for manifestfil fra tekst (*.txt) til BED (*.bed) i instruktioner i oprettelse af en kørsel. Rettet konsensus-VCF-filer til VCF-filer i afsnittet analyseoutput.
200014776 v01	August 2022	Tilføjelse: Afsnittet Indstillinger. Afsnittet Filtrering af systematisk støj. Opdateret anvisninger i oprettelse af kørsler til inkludere mere detaljerede oplysninger. Rettet stavefejl og grammatiske fejl. Angivet at anvisningerne er beregnet til programmet, når det bruges med NovaSeq 6000Dx-instrumentet. Opdateret oplysninger om indholdet af VCF-outputfiler.
200014776 v00	Marts 2022	Oprindelig udgivelse.

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 kommunikerer, 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.

Instruktionerne 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 INSTRUKTIONERNE 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).

© 2022 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](http://www.illumina.com/company/legal.html).

# Indholdsfortegnelse

Revisionshistorik .....	ii
Oversigt .....	1
Analysemetoder .....	1
Oprettelse af en kørsel .....	4
Indstillinger .....	6
Analyseoutput .....	7
FASTQ-filer .....	8
BAM-filer .....	8
VCF-filer .....	9
Visning af analyseresultater .....	14
<b>Teknisk hjælp .....</b>	<b>15</b>

# Oversigt

DRAGEN™ for Illumina® DNA Prep with Enrichment Dx-programmet udfører demultipleksring, FASTQ-generering, læsningskortlægning og alignment til et referencegenom og variantbestemmelse afhængigt af den valgte analysearbejdsgang.

## Analysemetoder

DRAGEN for Illumina DNA Prep with Enrichment Dx udfører demultipleksring, FASTQ-generering, læsningskortlægning og alignment til et referencegenom afhængigt af den valgte arbejdsgang:

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

## 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. flowcellelinje, pr. sekventeringslæsning for hver prøve. Prøvens navn, som angives under kørsels konfiguration, er inkluderet i FASTQ-filens navn. FASTQ-filer er det primære alignment-input. Det første trin i FASTQ-generering er demultipleksring. Demultipleksring tildeler clustre, der passerer filteret, til en prøve, ved at sammenligne hvert indeksslæsningssekvens med de indekssekvenser, der er angivet for kørslen. Ingen kvalitetsværdier tages i betragtning på dette trin. Indeksslæsninger 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 indekssekvensen matcher præcist, eller når der er højst én uoverensstemmelse pr. indeksslæsning.

Software inkluderer ORA-komprimering til at komprimere FASTQ-filer. Ved brug af ORA-formatet (\*.ora) bevares md5-kontrolsummen for FASTQ-indholdet efter en komprimerings- og dekomprimeringscyklus for at sikre tabsfri komprimering.

## Kortlægning og alignment af DNA

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-alignment på placeringerne med den højeste densitet af kimenmatch. 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 alignments mellem

læsningen og referencen. Smith-Waterman genererer scorer for hver af disse kandidat-alignment-placeringer, der bruges til at evaluere om den bedste alignment, der passerer gennem denne matrixcelle, når den via en nuklotidmatch eller mismatch (diagonal bevægelse), en deletion (vandret bevægelse) eller en insertion (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 alignment, der vælges.

De specifikke værdier, der vælges til scorer i denne algoritme, indikerer hvordan man balancerer, i tilfælde af en alignment med flere mulige tolkninger, muligheden for en indel, modsat en eller flere SNP'er, eller om der foretrækkes en alignment uden klipning. DRAGENs standard scoringsværdier er rimelige til at tilpasse læsninger af moderat længde til et helt humant referencegenom til variantbestemmelsesprogrammer. Alle sæt af Smith-Waterman-scoringsparametre repræsenterer en upræcis model af genmutation og sekventeringsfejl. Anderledes justerede alignment-scoringsværdier kan være mere passende for nogle programmer.

## DRAGEN bestemmelse af kimcellevarianter

DRAGEN Germline Small Variant Caller tager kortlagte og tilpassede DNA-læsninger som input og bestemmer SNP'er og indeler gennem en kombination af kolonnevis påvisning og lokal *de novo*-samling af haplotyper.

De referenceområder, der kan bestemmes, identificeres først med tilstrækkelig alignment-dækning. Inden for disse referenceområder vil en hurtig scanning af de sorterede læsninger identificere aktive områder, som er centreret rundt 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 modtage 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 backbone. 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 et par skjult Markov-model (HMM).

Ved scanning efter referenceplacering over det aktive område dannes der kandidat-genotyper 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.

I gVCF-tilstanden, der bruges til skalerbar variantbestemmelse af multiple prøver, kan DRAGEN Germline Small Variant Caller køres pr. prøve for at generere en midlertidig genomvariantbestemmelsesfil (gVCF). gVCF kan derefter bruges til effektiv, fælles genotypebestemmelse af multiple prøver, hvilket muliggør den hurtige, gradvise behandling af prøverne og skaleringen til store kohortestørrelser.

Da DRAGEN Germline Small Variant Caller har algoritmer, hvormed den effektivt kan skelne mellem korrelerede fejl og sande varianter, er filtreringsreglerne meget enkle.

## DRAGEN bestemmelse af somatiske varianter

DRAGEN Somatic Small Variant Caller tager kortlagte og tilpassede DNA-læsninger som input og bestemmer SNV'er og indeler gennem lokal *de novo*-samling af haplotyper i det aktive område.

De referenceområder, der kan bestemmes, identificeres først med tilstrækkelig alignment-dækning. Inden for disse referenceområder vil en scanning af de sorterede læsninger identificere aktive områder, som er centreret rundt om samlingen af kolonner med evidens for en variant i tumorklæsningerne. 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 backbone. 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 TLOD-scoren vil DRAGEN Somatic Small Variant Caller først scanne efter referenceplacering for hver kandidat for somatisk hændelse samt referencehændelsen over det aktive område. Den betingede sandsynlighed  $P(r|e)$  for at observere hver overlappende læsning estimeres 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 evidensen for, at en ALT-allel er til stede i tumorprøven ved en given locus.

# Oprettelse af en kørsel

Brug følgende trin for at oprette en kørsel i Illumina Run Manager enten på NovaSeq 6000Dx eller ved brug af en browser på en netværksforbundet computer. Prøvedata kan indlæses manuelt eller ved at importere et prøveark.

## Program- og kørselsindstillinger

1. Vælg **Create Run** (Opret kørsel) på skærmen Runs (Kørsler).
2. Vælg DRAGEN for Illumina DNA Prep with Enrichment Dx-programmet, og vælg så **Next** (Næste).
3. Indlæs et kørselsnavn på skærmen Run Settings (Kørselsindstillinger). Kørselsnavnet identificerer kørslen fra sekventering gennem analysen.
4. **[Valgfrit]** Angiv en kørselsbeskrivelse for yderligere at identificere kørslen.
5. Kontrollér, at det bibliotekklargøringsæt, der er valgt, er et Illumina DNA Prep with Enrichment Dx-biblioteksklargøringsæt.
6. Vælg det ønskede indeksadaptersæt.
7. Angiv Read Length (Læsningslængden).  
Read 1 (Læsning 1) og Read 2 (Læsning 2) har en standardværdig på 151 cyklusser.  
Index 1 (Indeks 1) og Index 2 (Indeks 2) har en fast værdi på 10 cyklusser.
8. **[Valgfrit]** Angiv et biblioteksror-ID.
9. Vælg **Next** (Næste).

## Prøvedata

Brug tabellen på skærmen Sample Data (Prøvedata) til manuelt at angive prøveoplysningerne. Eller vælg **Import Samples** (Importér prøver) for at uploade prøveoplysningerne. Se [Importér prøver på side 5](#) for oplysninger om importering af prøveoplysninger.

### Indtast prøver manuelt

1. Angiv et unikt prøve-ID i feltet Sample ID (Prøve-ID).
2. Brug **Plate - Well Position** (Plade - brøndplacering) til at vælge brøndplaceringen.  
Felterne i7 Index, Index 1, i5 Index og Index 2 udfyldes automatisk.
3. **[Valgfrit]** Angiv et biblioteksnavn.
4. Tilføj rækker, og gentag trin 1–3 efter behov, indtil alle prøverne er føjet til tabellen.
5. Vælg **Next** (Næste).



## Importér prøver

Der kan downloades en skabelon (\*.csv) fra skærmen Sample Data (Prøvedata), når der planlægges en kørsel i Illumina Run Manager via en browser på en netværksforbundet computer.

1. Vælg **Download Template** (Download skabelon) for at downloade en ikke-udfyldt CSV-fil.
2. Åbn CSV-filen, indtast prøveoplysningerne, og gem filen.  
Prøveark-CSV-filen inkluderer følgende datakolonner: Sample ID (Prøve-ID), Plate - Well Position (Plade - brøndplacering), **Optional** Library Name (Valgfrit Biblioteksnavn).
3. Vælg **Import Samples** (Importér prøver), og gå til placeringen for CSV-filen.
4. Vælg **Next** (Næste).

## Analyseindstillinger

1. Vælg den ønskede analysearbejdsgang:
  - FASTQ-generering
  - Kimcelle FASTQ- og VCF-generering for en kimcellerarbejdsgang
  - Somatisk FASTQ- og VCF-generering for en somatisk arbejdsang
2. **[Valgfrit]** Vælg om ønsket afkrydsningsfeltet **Generate ORA compressed FASTQs** (Generér ORA-komprimerede FASTQ'er) for at aktivere FASTQ ORA-komprimering.
3. **[Arbejdsgange for VCF-generering]** Brug rullemenuen **Manifest File Selection** (Valg af manifestfil) for at vælge en manifestfil.  
En manifestfil er påkrævet input til DRAGEN for Illumina DNA Prep with Enrichment Dx. Manifestfilen er en tabulatorsepareret BED (\*.bed), der angiver navnene og placeringerne på de målrettede referenceområder.
4. **[Arbejdsgang for somatisk FASTQ- og VCF-generering]** Brug rullemenuen **Noise File Selection** (Valg af støjfil) for at vælge en støjfil.  
Der kan angives en BED-fil med centerspecifikt støjniveau til at frasortere systematisk støj. Se [Støjfiltrering på side 6](#) for yderligere oplysninger.
5. Vælg **Next** (Næste).

## Kørsels- oversigt

1. På skærmen Review (Oversigt) skal du gennemgå de oplysninger, der er angivet på skærbillederne Run Settings (Kørselsindstillinger), Sample Data (Prøvedata) og Analysis Settings (Analyseindstillinger).
2. Vælg **Save** (Gem).  
Kørslen gemmes under fanen Planned (Planlagt) på skærmen Runds (Kørsler).

## Indstillinger

Vælg programmet på skærmen Applications (Programmer) for at se de aktuelle indstillinger og ændre indstillingerne.

### Konfiguration

Konfigurationsskærmen viser følgende programindstillinger:

- **Library Prep Kits** (Biblioteksklarkøringssæt) - Viser standard biblioteksklarkøringssættet for programmet. Denne indstilling kan ikke ændres.
- **Index Adapter Kits** (Indeksadaptersæt) - Viser standard indeksadaptersættet 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 radioknappen **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 for appen.

### Støjfiltrering

Filtrering af systematisk støj er tilgængelig ved brug af den somatiske arbejdsgang. Filteret kan bruges i tilstanden Tumor-Normal (Tumor-normal), men er især praktisk anvendeligt i kørslerne Tumor-Only (Kun tumor), hvor en matched normal ikke er tilgængelig.

BED-filen for systematisk støj skal genereres fra normale prøver. Det anbefales at opbygge filer for systematisk støj, der er specifikke for biblioteksklarkøring, sekventeringssystem og panel. Det anbefales, at der bruges ca. 50 normale prøver til generering af støjfilen.

# Analyseoutput

DRAGEN for Illumina DNA Prep with Enrichment Dx gemmer følgende oplysninger i analysemappen. Det er kun kimcellerarbejdsgange og somatiske arbejdsgange, der genererer en PDF.

- 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

Programmet vil generere følgende outputfiler. De eksakte, genererede filer vil afhænge af hvilken analysearbejdsgang, der bruges. Outputfilerne findes i analysemappen.

Outputfil	Beskrivelse
FASTQ (*.fastq.gz eller *.fastq.ora)	Midlertidige filer med kvalitetsscorede basebestemmelser. FASTQ-filer er det primære input til alignment-trinnet. Hvis der vælges ORA-komprimering, vil filnavnet afspejle dette.
Alignment-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 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
- Et plustegn (+)
- Phred-kvalitetsscorerne i et ASCII + 33-kodet format

Prøveidentifikatoren er formateret som følger.

```
@Instrument:KørselsID:FlowCelleID:Bane:Flise:X:Y
LæsningsNum:FilterFlag:0:PrøveNummer
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 (SAM, 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 sidehovedsektion og en alignment-sektion:

- Sidehoved - Indeholder oplysninger om hele filen, såsom prøvenavn, prøvelængde og alignment-metode. Alignments i alignment-sektionen er forbundet med specifikke oplysninger i sidehovedsektionen.
- Alignments - Indeholder læsningsnavn, læsningssekvens, læsningskvalitet, alignment-oplysninger og brugerdefinerede tags. Læsningsnavnet inkluderer kromosomet, startkoordinaten, alignment-kvaliteten og den matchede beskrivelsesstreng.

Alignment-sektionen inkluderer følgende oplysninger for hver læsning eller læsningspar:

- AS: Paired-end alignment-kvalitet.
- 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: Enkeltende-alignment-kvalitet.

- XC: Matchbeskrivelsesstreng.
- XN: Amplikon-navnemærke, der registrerer det amplikons-ID, der er forbundet med læsningens BAM-indeksfiler (\*.bam.bai), giver et indeks over den tilsvarende BAM-fil.

## VCF-filer

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

VCF-filens sidehoved inkluderer VCF-filformatets version og variantbestemmelsesprogrammets version og en liste over de annoteringer, der er brugt i resten af filen. VCF-sidehovedet inkluderer også referencegenomfilen og BAM-filen. Den sidste linje i sidehovedet 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 deletion 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 insertion 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-filannoteringer

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

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

Overskrift	Beskrivelse
INFO	<p>Kimcellarbejdsgang, mulige INFO-poster inkluderer:</p> <ul style="list-style-type: none"> <li>• <b>AC</b> - Alleltal i genotyper for hver ALT-allel, i samme rækkefølge som angivet.</li> <li>• <b>AF</b> - Allelfrekvens for hver ALT-allel, i samme rækkefølge som angivet.</li> <li>• <b>AN</b> - Samlet antal alleler i bestemte genotyper.</li> <li>• <b>DB</b> - dbSNP-medlemsskab.</li> <li>• <b>FS</b> - P-værdi på Phred-skala ved brug af Fishers eksakte test til påvisning af bias af strenge.</li> <li>• <b>QD</b> - Variantkonfidens/-kvalitet efter dybde.</li> <li>• <b>R2_5P_bias</b> - Score baseret på bias af makker og afstand fra 5 prim-ende.</li> <li>• <b>SOR</b> - Symmetrisk odds ratio for 2x2 kontingenstabel til påvisning af bias af strenge.</li> <li>• <b>DP</b> - Omtrentlig læsningsdybde (informativ og ikke-informativ). Nogle læsninger kan være blevet filtreret på mapq osv.</li> <li>• <b>END</b> - Intervallets stopposition.</li> <li>• <b>FractionInformativeReads</b> - Fraktionen af informative læsninger ud af det samlede antal læsninger.</li> <li>• <b>MQ</b> - RMS-kortlægningskvalitet.</li> <li>• <b>MQRankSum</b> - Z-score fra Wilcoxons rank-sum test af kvaliteten af Alt- vs. Ref-læsningskortlægninger.</li> <li>• <b>ReadPosRankSum</b> - Z-score fra Wilcoxons rank-sum test af bias af Alt vs. Ref-læsningsplaceringer.</li> <li>• <b>SOMATIC</b> - Mindst en variant på denne placering er somatisk.</li> </ul> <p>Somatisk arbejdsgang, mulige INFO-poster inkluderer:</p> <ul style="list-style-type: none"> <li>• <b>DP</b> - Omtrentlig læsningsdybde (informativ og ikke-informativ). Nogle læsninger kan være blevet filtreret på mapq osv.</li> <li>• <b>END</b> - Intervallets stopposition.</li> <li>• <b>FractionInformativeReads</b> - Fraktionen af informative læsninger ud af det samlede antal læsninger.</li> <li>• <b>MQ</b> - RMS-kortlægningskvalitet.</li> <li>• <b>MQRankSum</b> - Z-score fra Wilcoxons rank-sum test af kvaliteten af Alt- vs. Ref-læsningskortlægninger.</li> <li>• <b>ReadPosRankSum</b> - Z-score fra Wilcoxons rank-sum test af bias af Alt vs. Ref-læsningsplaceringer.</li> <li>• <b>AQ</b> - Score for systematisk støj.</li> <li>• <b>hotspot</b> - Kendt somatisk sted, brugt til at øge tilliden til bestemmelsen.</li> <li>• <b>SOMATIC</b> - Mindst en variant på denne placering er somatisk.</li> </ul>



Overskrift	Beskrivelse
FORMAT	<p>Formatkolonnen har kolonseparerede felter. For eksempel GT:GQ.</p> <p>De mulige felter kimcellerarbejdsgangen inkluderer:</p> <ul style="list-style-type: none"> <li>• <b>AD</b> - Allelgybder (tæller kun informative læsninger ud af det samlede antal læsninger) for ref- og alt-allelerne i den rækkefølge, de er angivet.</li> <li>• <b>AF</b> - Allelfraktioner for alt-alleler i den rækkefølge, de er angivet.</li> <li>• <b>DP</b> - Omtrentlig læsningsdybde (læsninger med MQ=255 eller med dårlige makker filtreres).</li> <li>• <b>F1R2</b> - Antal læsninger i F1R2-parretning, der understøtter hver allel.</li> <li>• <b>F2R1</b> - Antal læsninger i F2R1-parretning, der understøtter hver allel.</li> <li>• <b>GP</b> - Posteriore sandsynligheder for genotyper ifølge Phred-skalaen, som defineret i VCF-specifikationen.</li> <li>• <b>GQ</b> - Genotypekvalitet.</li> <li>• <b>GT</b> - 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.</li> <li>• <b>MB</b> - Statistik pr. prøvekomponent til påvisning af makker-bias.</li> <li>• <b>PL</b> - Normaliserede sandsynligheder for genotyper ifølge Phred-skalaen, som defineret i VCF-specifikationen.</li> <li>• <b>PRI</b> - Tidligere sandsynligheder for genotyper ifølge Phred-skalaen.</li> <li>• <b>PS</b> - ID-oplysninger for fysisk faseinddeling, hvor hver unikt ID i en given prøve (men ikke på tværs af prøver) forbinder poster inden for en faseinddelingsgruppe.</li> <li>• <b>SB</b> - Statistik pr. prøvekomponent, som består af Fishers eksakte test til påvisning af streng-bias.</li> <li>• <b>SQ</b> - Somatisk kvalitet.</li> </ul> <p>De tilgængelige felter for somatisk arbejdsgang inkluderer:</p> <ul style="list-style-type: none"> <li>• <b>AD</b> - Allelgybder (tæller kun informative læsninger ud af det samlede antal læsninger) for ref- og alt-allelerne i den rækkefølge, de er angivet.</li> <li>• <b>AF</b> - Allelfraktioner for alt-alleler i den rækkefølge, de er angivet.</li> <li>• <b>DP</b> - Omtrentlig læsningsdybde (læsninger med MQ=255 eller med dårlige makker filtreres).</li> <li>• <b>F1R2</b> - Antal læsninger i F1R2-parretning, der understøtter hver allel.</li> <li>• <b>F2R1</b> - Antal læsninger i F2R1-parretning, der understøtter hver allel.</li> <li>• <b>GT</b> - 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.</li> <li>• <b>MB</b> - Statistik pr. prøvekomponent til påvisning af makker-bias.</li> <li>• <b>PS</b> - ID-oplysninger for fysisk faseinddeling, hvor hver unikt ID i en given prøve (men ikke på tværs af prøver) forbinder poster inden for en faseinddelingsgruppe.</li> <li>• <b>SB</b> - Statistik pr. prøvekomponent, som består af Fishers eksakte test til påvisning af streng-bias.</li> <li>• <b>SQ</b> - Somatisk kvalitet.</li> </ul>

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

## Genom VCF-filer

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

# Visning af analyseresultater

Kørsler der aktuelt er i gang vises under fanen Active (Aktiv). Fuldførte kørsler vises under fanen Completed (Fuldført). Se [NovaSeq 6000Dx Product Documentation \(Produktdokumentation til NovaSeq 6000Dx\)](#) (dokumentnr. 200010105) for yderligere oplysninger om visning af resultater.

# Teknisk hjælp

Kontakt Illumina teknisk support for at få teknisk hjælp.

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

## Telefonnumre til Illumina teknisk support

Område	Gratis	Internationalt
Australien	+61 1800 775 688	
Østrig	+43 800 006249	+43 1 9286540
Belgien	+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
Frankrig	+33 8 05 10 21 93	+33 1 70 77 04 46
Tyskland	+49 800 101 4940	+49 89 3803 5677
Hong Kong, Kina	+852 800 960 230	
Indien	+91 8006500375	
Indonesien		0078036510048
Irland	+353 1800 936608	+353 1 695 0506
Italien	+39 800 985513	+39 236003759
Japan	+81 0800 111 5011	
Malaysia	+60 1800 80 6789	
Holland	+31 800 022 2493	+31 20 713 2960
New Zealand	+64 800 451 650	
Norge	+47 800 16 836	+47 21 93 96 93
Filippinerne	+63 180016510798	
Singapore	1 800 5792 745	
Sydkorea	+82 80 234 5300	

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

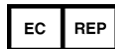
**Sikkerhedsdatablade (SDS'er)** – Kan findes på Illuminas websted på [support.illumina.com/sds.html](https://support.illumina.com/sds.html).

**Produktdokumentation** – Kan downloades på [support.illumina.com](https://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

CE



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

**Australsk sponsor**

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

TIL IN VITRO-DIAGNOSTISK BRUG

© 2022 Illumina, Inc. Alle rettigheder forbeholdes.

**illumina**<sup>®</sup>