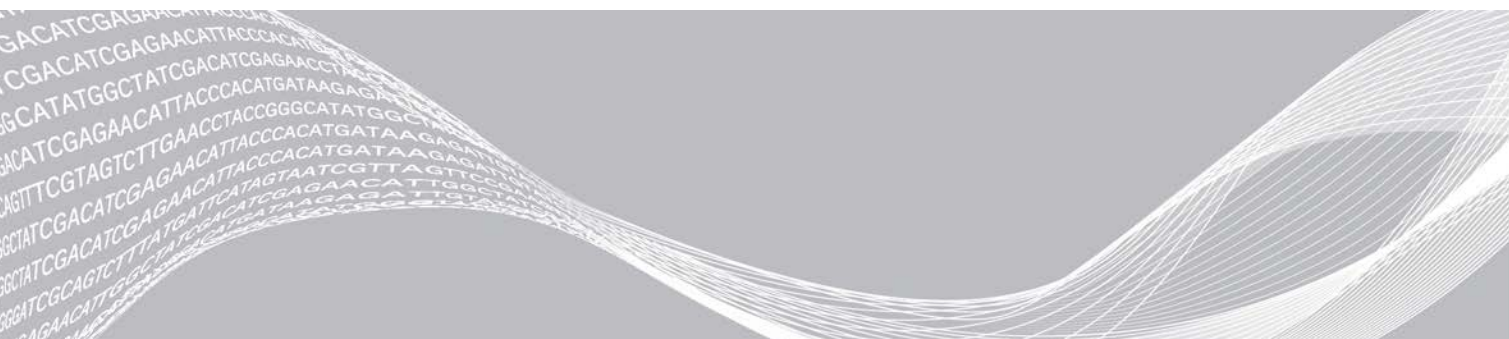


# Local Run Manager Somatic Variant-analysemodulet

Arbejdsprocesvejledning for NextSeq 550Dx

KUN TIL IN VITRO-DIAGNOSTIK

Oversigt	3
Indtastning af kørselsoplysninger	4
Analysemetoder	6
Visning af kørsels- og prøvedata	7
Analysis Report (Analyserapport)	7
Analyseoutputfiler	9
Basebestemmelse og indeksdiversitet	15
Revisionshistorik	17
Teknisk hjælp	18



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 kommunikeres, 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).

© 2021 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).

## Oversigt

Local Run Manager Somatic Variant-modulet er beregnet til brug sammen med Illumina TruSeq Custom Amplicon Kit Dx-analysen og NextSeq 550Dx. Når analysen bruges sammen med Somatic Variant-modulet, er den beregnet til at klargøre biblioteker, der bruges til sekventering af DNA fra formalinfikseret, paraffinindstøbt (FFPE) væv. Analysen detekterer somatiske mutationer ved lave variantfrekvenser.

Analysemodulet vurderer korte regioner af forstærket DNA, eller amplicons, til varianter. Fokuseret sekventering af amplicons muliggør høj dækning af særlige regioner på tværs af et stort antal prøver. Analysemodulet udfører sekundær analyse og rapportgenerering ud fra sekventeringskørsler med en dual strand-tilgang, der omfatter forward og reverse oligo pools. Se indlægssedlen *TruSeq Custom Amplicon Kit Dx* (dokumentnr. 1000000029772).

Somatic Variant-analysemodulet kræver 300-cyklus sekventeringsforbrugsartikler. Du kan finde flere oplysninger på indlægssedlen til *NextSeq 550Dx High Output Reagent Kit v2* eller *NextSeq 550Dx High Output Reagent Kit v2.5*.

## Om denne vejledning

Denne vejledning indeholder instruktioner om konfiguration af kørselsparametre for sekventering og analyse i Somatic Variant-analysemodulet. Vedrørende oplysninger om Local Run Manager-dashboardet og systemindstillingerne henvises til *NextSeq 550Dx Instrumentreferencevejledning* (dokumentnr. 1000000009513).

## Visning af Local Run Manager

Local Run Manager-interfacet vises inde i NextSeq 550Dx Operating Software (NOS) eller via en webbrowser. Den understøttede webbrowser er Chromium.



### BEMÆRK!

Hvis du bruger en browser, der ikke understøttes, kan du downloade den understøttede browser, når du får meddelelsen "Confirm Unsupported Browser" (Bekræft ikke-understøttet browser). Vælg "**here**" (her) for at downloade den understøttede version af Chromium.

## Visning på instrumentskærmen

- 1 Vælg en af følgende muligheder for at få vist brugergrænsefladen til Local Run Manager på instrumentskærmen:
  - ▶ Vælg **Local Run Manager** på NOS-startsiden.  
Klik på X'et i øverste højre hjørne for at vende tilbage til NOS, når du er færdig.
  - ▶ Vælg ikonet Minimer NOS, åbn Chromium-browseren på instrumentet, og skriv **http://localhost** i adresselinjen.  
Det er kun administratorbrugere, der kan minimere NOS.

## Visning på en netværkscomputer

- 1 Åbn Chromium-browseren på en computer, der er forbundet til det samme netværk som instrumentet, og opret forbindelse ved hjælp af instrumentets IP-adresse eller navn. For eksempel **http://mitinstrument**.

## Indtastning af kørselsoplysninger

### Konfiguration af parametre

- 1 Log ind på Local Run Manager.
- 2 Vælg **Create Run** (Opret kørsel), og vælg **Somatic Variant**.
- 3 Indtast et kørselsnavn, der identificerer kørslen fra sekventering via analyse.  
Brug alfanumeriske tegn, mellemrum, understregninger, binde- eller tankestreger.
- 4 **[Valgfri]** Indtast en kørselsbeskrivelse, der gør det nemmere at identificere kørslen.  
Brug alfanumeriske tegn, mellemrum, understregninger, binde- eller tankestreger.
- 5 Vælg antallet af prøver og indekssættet på rullelisten.  
Tag højde for følgende oplysninger, når du vælger.
  - ▶ Rullelisten indeholder antallet af prøver i et indekssæt. "24-Set 1" angiver f.eks. 24 prøver, der skal testes, med indekser fra indekssæt 1.
  - ▶ Indekssætnumre henviser til forskellige sæt af i5-indekser. Både sæt 1 og sæt 2 giver indeksdiversitet. To indekssæt tilbydes for at forebygge udtømming af et enkelt sæt.
  - ▶ Vælg det antal prøver, der er tættest på det antal prøver, du tester. Hvis det præcise antal prøver ikke findes på listen, vælges det nærmeste antal, men under det antal du tester. Hvis du f.eks. vil teste 18 prøver, skal du vælge 16 prøver.
  - ▶ Prøvebrønde og indeksskcombinationer, der opfylder kravene til indeksdiversitet fremhæves med grønt. Hvis du vælger andre brønde og indeksskcombinationer, når du gemmer kørslen, får du en meddelelse, hvis kravene til indeksdiversitet ikke er opfyldt.

### Import af manifestfiler for kørslen

- 1 Sørg for, at de manifeste, som du ønsker at importere, ligger på en tilgængelig netværksplacering eller på et USB-drev.
- 2 Vælg **Import Manifests** (Importér manifeste).
- 3 Gå til manifestfilen, og vælg de manifeste, du vil tilføje.



#### **BEMÆRK!**

For at gøre manifestfiler tilgængelige for alle kørsler med Somatic Variant-analysemodul, skal manifeste tilføjes ved hjælp af Module Settings-funktionen. Denne funktion kræver administratorrettigheder. For yderligere oplysninger henvises til *NextSeq 550Dx Instrumentreferencevejledning (dokumentnr. 1000000009513)*.


### Angivelse af prøver til kørslen

Angiv prøverne til kørslen ved hjælp af et af alternativerne og følgende vejledning.


- ▶ **Indtast prøverne manuelt** – Brug den tomme tabel på skærmen Create Run (Opret kørsel).
- ▶ **Importér prøver** – Gå til en ekstern fil i et kommasepareret værdiformat (\*.csv). Du kan downloade en skabelon på skærmen Create Run (Opret kørsel).

Når du har udfyldt prøvetabellen, kan du eksportere prøveoplysningerne til en ekstern fil. Brug filen som reference, når du forbereder biblioteker, eller importér filen til endnu en kørsel.

## Manuel indtastning af prøver

- 1 Indtast et unikt prøvenavn i feltet Sample Name (Prøvenavn).  
Brug alfanumeriske tegn, binde- eller tankestreger eller understregninger.  
Prøvenavnet udfyldes automatisk i den tilsvarende brønd i den anden pool.
- 2 **[Valgfri]** Højreklik ved positive eller negative kontrolprøver, og vælg kontroltype.  
Kontrollen i én prøvebrønd udfylder automatisk den tilsvarende brønd i den anden pulje med samme kontrol.
- 3 **[Valgfrit]** Indtast en prøvebeskrivelse i feltet Sample Description (Prøvebeskrivelse).  
Brug alfanumeriske tegn, binde- eller tankestreger eller understregninger.  
Prøvebeskrivelsen udfyldes automatisk i den tilsvarende brønd i den anden pool.  
Prøvebeskrivelser er forbundet med et prøve-ID. Prøvebeskrivelser overskrives, hvis det samme prøve-ID bruges igen i en senere kørsel.
- 4 Vælg en Index 1-adapter på rullelisten Index 1 (i7).  
Når du anvender de foreslåede prøvebrønde, udfylder softwaren automatisk i7 og i5-indeksadaptere, der overholder kravene til indeksdiversitet. Hvis det præcise antal prøver, du tester, ikke findes på listen, skal du huske at vælge indeksadaptere til ekstra brønde. Hvis du skal vælge indekser til ekstra brønde, eller hvis du ikke bruger de anbefalede indeksadapterkombinationer, før du vælger, skal du huske at læse *Basebestemmelse og indeksdiversitet på side 15*.
- 5 Vælg en Index 2-adapter fra rullelisten Index 2 (i5).
- 6 Vælg en manifestfil på rullelisten Manifest.  
Prøver i pulje A kræver et anden manifest end prøverne i pulje B.
- 7 Du kan gennemse, udskrive eller gemme pladelayoutet som reference til klargøring af biblioteker:
  - ▶ Vælg ikonet  **Print** (Udskriv) for at vise pladelayoutet. Vælg **Print** (Udskriv) for at udskrive pladelayoutet.
  - ▶ Vælg **Export** (Eksportér) for at eksportere prøveoplysninger til en ekstern fil.  
Kontrollér, at manifest- og prøveoplysninger er korrekte. Forkerte oplysninger kan påvirke resultater.
- 8 Vælg **Save Run** (Gem kørsel).

## Import af prøver

- 1 Vælg **Import Samples** (Importér prøver), og gå til prøveoplysningsfilens placering. Der er to typer filer, du kan importere.
  - ▶ Vælg **Template** (Skabelon) på skærmen Create Run (Opret kørsel) for at lave en ny pladelayout. Skabelonfilen indeholder de rigtige kolonneoverskrift til import. Indtast prøveoplysninger i hver kolonne om prøverne i kørslen. Slet eksempeloplysninger i ubrugte celler, og gem derefter filen.
  - ▶ Brug en fil med prøveoplysninger, der blev eksporteret fra Somatic Variant-modulet ved hjælp af funktionen Export (Eksportér).
- 2 Vælg ikonet  **Print** (Udskriv) for at vise pladelayoutet.
- 3 Vælg **Print** (Udskriv) for at udskrive pladelayoutet som reference til klargøring af biblioteker.
- 4 **[Valgfri]** Vælg **Export** (Eksportér) for at eksportere prøveoplysninger til en ekstern fil.  
Kontrollér, at manifest- og prøveoplysninger er korrekte. Forkerte oplysninger kan påvirke resultater.
- 5 Vælg **Save Run** (Gem kørsel).

## Redigering af en kørsel

Vedrørende instruktioner om redigering af informationen i din kørsel før sekventering henvises til *NextSeq 550Dx Instrumentreferencevejledning (dokumentnr. 1000000009513)*.

## Analysemetoder

Somatic Variant-analysemodulet udfører følgende analysetrin og skriver derefter analyseoutputfiler til Alignment-mappen.

- ▶ Demultiplekserer indekslæsninger
- ▶ Genererer FASTQ-filer
- ▶ Justerer til en reference
- ▶ Identificerer varianter

## Demultipleksering

Demultiplekseringen sammenligner hver indekslæsningssekvens med de indeksskvenser, der er angivet for kørslen. Ingen kvalitetsværdier tages i betragtning på dette trin.

Indekslæsninger identificeres ved hjælp af følgende trin:

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

## Generering af FASTQ-fil

Efter demultipleksering genererer softwaren intermediære analysefiler i FASTQ-formatet, der er et tekstformat, som bruges til at repræsentere sekvenser. FASTQ-filerne indeholder læsninger for hver prøve med tilhørende kvalitetsgradueringer. Klynger, som ikke passerede filtret, medtages ikke.

Hver FASTQ-fil indeholder kun læsninger for én prøve, og prøvens navn indgår i navnet på FASTQ-filen. FASTQ-filer er det primære alignmentinput. Der genereres otte FASTQ-filer pr. prøve pr. oligo pool, fire fra Læsning 1 og fire fra Læsning 2, der giver i alt 16 FASTQ-filer pr. prøve.

## Tilpasning

Under alignmenttrinnet aligner den bandede Smith-Waterman-algoritme klynger fra hver prøve i forhold til ampliconsekvenser, der er specificeret i manifestfilen.

Smith-Waterman algoritmen udfører semiglobale tilpasninger af sekvenser for at klarlægge ens områder mellem to sekvenser. I stedet for at sammenligne hele sekvensen sammenligner Smith-Waterman algoritmen segmenter af alle mulige længder.

Hver læsning med parret afslutning analyseres med hensyn til læsningens tilpasning med de relevante sondesekvenser for den pågældende læsning.

- ▶ Læsning 1 analyseres i forhold til den omvendte komplement af den senere DLSO (Downstreams Locus-Specific Oligos).
- ▶ Læsning 2 vurderes i forhold til ULSO (Upstream Locus-Specific Oligos).

- ▶ Hvis starten af en læsning matcher en probesekvens med højst tre forskelle (mis-matches eller skift på grund af ledende indels), justeres læsningen i hele sin længde i forhold til ampliconmålet for den pågældende sekvens.
- ▶ På grund af prøvekemien observeres indels inden for DLSO og ULSO ikke.

Tilpasninger filtreres fra tilpasningsresultaterne på baggrund af uoverensstemmelsesrater, enten over interesseområdet eller hele ampliconet, afhængigt af ampliconlængden. De frafilterede tilpasninger skrives i tilpasningsfilerne som ikke-tilpassede og anvendes ikke til variantbestemmelsen.

## Variantbestemmelse



Pisces Variant Caller er udviklet af Illumina og identificerer varianter, der er til stede ved lav frekvens i DNA-prøven.

Pisces Variant Caller identificerer SNV'er, MNV'er og små indels i tre trin:

- ▶ Overvejer separat hver position i referencegenomet
- ▶ Tæller baser på den givne position for alignede læsninger, der overlapper positionen
- ▶ Beregner en variantscore, der måler bestemmelsens kvalitet ved hjælp af Poisson-modellen. Varianter med en kvalitetsscore under Q30 udelukkes.

Varianter bestemmes først separat for hver pool. Derefter sammenlignes varianter fra hver pool og kombineres til en enkelt outputfil. Hvis en variant findes i begge pools og passerer alle filtre, der er angivet i *VCF-filnoter på side 12*, markeres varianten som PASS i variantbestemmelsesfilen (VCF).

## Visning af kørsels- og prøvedata

- 1 Klik på kørselsnavnet på Local Run Manager-dashboardet.
- 2 Gennemgå sekventeringskørselsens metrik på fanen Run Overview (Kørselsoversigt).
- 3 **[Valgfri]** Klik på ikonet **Copy to Clipboard**  (Kopier til skrivebord) for at kopiere outputkørselsmappens stinavn.
- 4 Klik på fanen Sequencing Information (Sekventeringsoplysninger) for at gennemgå kørselsparametrene og oplysningerne om forbrugsartikler
- 5 Klik på fanen Samples and Results (Prøver og resultater) for at se analyserapportens placering.
  - ▶ Hvis analysen blev gentaget, udvides rullelisten Select Analysis (Vælg analyse). Vælg så den relevante analyse.
- 6 Klik på ikonet **Copy to Clipboard**  (Kopier til skrivebord) for at kopiere analysemappens stinavn.

For yderligere oplysninger om fanerne Run Overview (Kørselsoversigt) og Sequencing Information (Sekventeringsoplysninger) samt om, hvordan analysen sættes i kø igen, henvises til *NextSeq 550Dx Instrument Reference Guide (document # 1000000009513)* (Oversigtsvejledning til NextSeq 550Dx-instrumentet (dokumentnr. 1000000009513)).

## Analysis Report (Analyserapport)

Analyserapporten opsummeres på fanen Samples and Results (Prøver og resultater) og som en samlet rapport i Alignmentmappen. Der findes også en rapport for hver prøve i PDF-format for hver prøve.

## Oplysningerne på fanerne Samples (Prøver) og Results (Resultater)

1 Klik på en prøve på listen for at se prøverapporten.

**Tabel 1 Kørsels- og prøveoplysninger**

Kolonneoverskrift	Beskrivelse
Run Status (status for sekventeringskørsel)	Angiver, om sekventeringskørslen er bestået eller dumpet.
Total Yield (GB) (Udbytte i alt (GB))	Antal baser, der bestemmes i sekventeringskørslen. Viser bestået grænse samt status som bestået eller dumpet.
% $\geq$ Q30	Procentdelen af læsninger i sekventeringskørslen med en kvalitetsscore på 30(Q30) eller derover. Viser bestået grænse samt status som bestået eller dumpet.
Sample Name (Prøvenavn)	Det prøvenavn, der blev givet, da kørslen blev oprettet.
Total PF Reads (PF-læsninger i alt)	Det samlede antal læsninger, der passerer filteret.
Read 1% $\geq$ Q30 (Læsning 1 % $\geq$ Q30)	Procentdelen af læsninger i Læsning 1 med en kvalitetsscore på 30(Q30) eller derover for prøven.
Read 2% $\geq$ Q30 (Læsning 2 % $\geq$ Q30)	Procentdelen af læsninger i Læsning 2 med en kvalitetsscore på 30(Q30) eller derover for prøven.
Autosomal Call Rate (Autosom bestemmelsesrate)	Antallet af genompositioner på tværs af autosomerne (kromosom 1 til 22), der opfylder en forud fastsat konfidensværdigrænse, divideret med det samlede antal forespurgte autosome genompositioner. Bestemmelsesraten beskrives pr. prøve og rapporteres som en procentdel, der beregnes som 1 minus (antal autosome positioner med ufuldstændige bestemmelser divideret med det samlede antal sekventerede autosome positioner.

**Tabel 2 Prøverapportoplysninger**

Kolonneoverskrift	Beskrivelse
Sample (Prøve)	Det prøvenavn, der blev givet, da kørslen blev oprettet.
Report Date (Rapportdato)	Den dato, hvor rapporten blev genereret.
Sample Information (Prøveoplysninger)	Det prøve-ID, der blev givet, da kørslen blev oprettet, læsninger i alt, der passede filteret i prøven, procentdelen af læsninger for prøven med en kvalitetsscore på 30 (Q30) eller derover, samt den autosome bestemmelsesrate.
Amplicon Summary (Ampliconoversigt)	Det samlede antal sekventerede ampliconregioner og den samlede længde i basepar af sekventerede amplicons i målregionerne for prøven i Pool A og Pool B og for manifestfilen, der bruges til hver pool. Manifestfilen specificerer referencegenomet og målreferenceregionerne, der bruges i alignmenttrinnet.
Read Level Statistics (Statistik på læsningsniveau)	Antal og procent af læsninger for prøven, der dækker hver position i referencen, for Læsning 1 og Læsning 2 i Pool A og Pool B.
Variants Summary (Variantoversigt)	Antal SNV'er, indsættelser og sletninger, der er detekteret for prøven, som passerede foreslåede værdier for at bestemme, om kvalitetsresultaterne er inden for et acceptabelt interval,
Coverage Summary (Dækningsoversigt)	Det samlede antal alignede baser divideret med målregionens størrelse, og procentdelen af ampliconregioner med dækningsværdier over den lave dækningsgrænse på 0,2 * amplicons middeldækning, for prøven i Pool A og Pool B.



Kolonneoverskrift	Beskrivelse
Coverage Plots (Dækningsplots)	Ampliconregionsplottets dækning viser dækningen på tværs af prøvens ampliconregioner. Regioner med dækningsværdier, der er lavere end dækningsgrænsen, er fremhævet med rødt. Gennemsnittet af alle værdier er angivet af en orange linje. Et plot er givet for dækningen af Pool A og Pool B.
Software Versions (Softwareversioner)	Softwareversioner, da prøven blev sekventeret. Omfatter NextSeq 550Dx Operating Software (NOS), Local Run Manager-software, RTA-software og Somatic Variant-modulversion.

## Analyseoutputfiler

Der genereres følgende analyseoutputfiler for Somatic Variant-analysemodulet, der giver analyseresultater for alignment og variantbestemmelse. Analyseoutputfiler placeres i Alignment-mappen.

Filnavn	Beskrivelse
Demultiplexering (*.txt)	Intermediære filer med oversigtsresultater af demultiplexering.
FASTQ (*.fastq.gz)	Intermediære filer med kvalitetsscorede basebestemmelser. FASTQ-filer er det primære input til alignmenttrinnet.
Alignmentfiler i BAM -formatet (*.bam)	Indeholder justerede læsninger for en given prøve.
Per-Pool variantbestemmelsesfiler i VCF -formatet (*.vcf)	Indeholder varianter, bestemt ved hver position fra enten forward pool eller reverse pool.
Variantbestemmelsesfiler i genom-VCF -formatet (*.genom.vcf.gz)	Indeholder genotypen for hver position, hvad enten den bestemmes som en variant eller som en reference.
Konsensusvariantbestemmelsesfiler i VCF -formatet (*.vcf.gz)	Indeholder varianter, bestemt ved hver position fra begge pools.
AmpliconCoverage_M1.tsv	Indeholder oplysninger om dækning pr. amplicon pr. prøve for hver leveret manifest. M# repræsenterer manifestnummeret.

## Demultiplexering af filformat

Demultiplexeringsprocessen læser den indekssækvens, der er forbundet til hver klynge, for at bestemme, hvilken prøve klyngen oprindeligt stammer fra. Mapping mellem klynger og prøvenummer skrives til en demultiplexeringsfil (\*.demux) for hver flise i flowcellen.

Navngivningsformatet for demultiplexeringsfilen er **s\_1\_X.demux**, hvor X er flisenummeret.

Demultiplexeringsfiler starter med en toptekst:

- ▶ Version (4 byte-heltal), p.t. 1
- ▶ Klyngeantal (4 byte-heltal)

Resten af filen består af prøveantal for hver klynge fra flisen.

Når demultiplexeringstrinnet er færdigt, genererer softwaren en demultiplexeringsfil med navnet **DemultiplexSummaryF1L1.txt**.

- ▶ I filnavnet repræsenterer **F1** flowcellenummeret.
- ▶ I filnavnet repræsenterer **L1** banenummeret.
- ▶ Demultiplexering giver en tabel med 1 række pr. flise og 1 kolonne pr. prøve, inkl. prøve 0.
- ▶ De mest almindeligt forekommende sekvenser i indeksslæsninger.

## FASTQ-filformat

FASTQ er et tekstbaseret filformat, der indeholder basebestemmelser og kvalitetsværdier for hver læsning. Hver post indeholder 4 linjer:

- ▶ Identifikatoren
- ▶ Sekvensen
- ▶ Et plustegn (+)
- ▶ Phred-kvalitetsscorerne i et ASCII + 33-kodet format

Identifikatoren formatteres som:

**@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-filformat

En BAM-fil (\*.bam) er den komprimerede binære version af en SAM-fil, der bruges til at repræsentere justerede sekvenser op til 128 MB. SAM-og BAM-formaterne er nærmere beskrevet i [samtools.github.io/hts-specs/SAMv1.pdf](https://samtools.github.io/hts-specs/SAMv1.pdf).

BAM-filer bruger filnavngivningsformatet **PrøveNavn\_S#.bam**, hvor # er prøvenummeret, der bestemmes af den rækkefølge, som prøverne er anført i for kørslen.

BAM-filer har et toptekstafsnit og et alignmentafsnit:

- ▶ **Header** (Toptekst) – Indeholder oplysninger om hele filen, såsom prøvenavn, prøvelængde og alignmentmetode. Alignments i alignmentafsnittet er forbundet med specifikke oplysninger i toptekstafsnittet.
- ▶ **Alignments** – Indeholder læst navn, læst sekvens, læst kvalitet, alignmentoplysninger og brugertilpassede tags. Det læste navn omfatter kromosomet, startkoordinaten, alignmentkvaliteten og matchdeskriptorstrengen.

Alignmentafsnittet omfatter følgende oplysninger for hver læsning eller læst par:

- ▶ **AS:** Paired-end-alignmentkvalitet.
- ▶ **BC:** Stregkodetag, der angiver det demultiplekserede prøve-ID, der er forbundet med læsningen.
- ▶ **SM:** Single-end alignmentkvalitet.
- ▶ **XC:** Match deskriptorstreng.
- ▶ **XN:** Amplicons navnetag, der registrerer det amplicon-ID, der er forbundet med læsningen.

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

## VCF-filformat

Variant Call Format (VCF) er et almindeligt filformat, der er udviklet af forskersamfundet inden for genomikken. Det indeholder oplysninger om varianter, der findes på bestemte positioner i et referencegenom. VCF-filer har endelsen .vcf.

VCF-filens toptekst omfatter VCF-filens formatversion og Variant Caller-version og angiver de noter, der bruges i resten af filen. VCF-topteksten omfatter også referencegenomfilen og BAM-filen. Den sidste linje i topteksten indeholder kolonneoverskrifterne til datalinjerne. Hver af datalinjerne i VCF-filen indeholder oplysninger om én variant.

## VCF-filtoptekster

Toptekst	Beskrivelse
CHROM	Referencegenomets kromosom. Kromosomer vises i samme rækkefølge som reference-FASTQ-filen.
POS	Variantens enkeltbaseposition i referencekromosomet. I forbindelse med SNP'er er denne position referencebasen med varianten. I forbindelse med indels eller sletninger er denne position referencebasen umiddelbart før varianten.
ID	Variantens rs-nummer, der stammer fra dbSNP.txt, hvis det er relevant. Hvis der er flere rs-numre på dette sted, er listen separeret med semikoloner. Hvis der ikke er nogen dbSNP-post på denne position, bliver der brugt en manglende værdimarkør ('.') .
REF	Referencegenotypen. F.eks. repræsenteres sletningen af et enkelt T som reference-TT og alternativ T. En A til T enkelt nukleotidvariant repræsenteres som reference A og alternativ T.
ALT	De alleler, der adskiller sig fra referencelæsningen. F.eks. repræsenteres en indsætning af et enkelt T som reference A og alternativ AT. En A til T enkelt nukleotidvariant repræsenteres som reference A og alternativ T.
QUAL	En Phred-skaleret kvalitetsscore, der tildeles af Variant Caller. Højere scorer angiver højere konfidens til varianten og lavere sandsynlighed for fejl. Ved en kvalitetsscore på Q er den estimerede fejlsandsynlighed $10^{-(Q/10)}$ . F.eks. har sættet af Q30-bestemmelser en fejlrate på 0,1 %. Mange Variant Callers tildeler kvalitetsscorer ud fra deres statistiske modeller, der er høje i forhold til den observerede fejlrate

## VCF-filnoter

Toptekst	Beskrivelse
<b>FILTER</b>	<p>Hvis alle filtre er passeret, skrives <b>PASS</b> i filterkolonnen.</p> <ul style="list-style-type: none"> <li>• <b>LowDP</b> – Gælder sites med en dækningsdybde under 450x i en pool. For ampliconpositioner, der dækkes af både forward og reverse læsningen, svarer dette til 900x enkeltlæsningsdækning.</li> <li>• <b>LowGQ</b> – Kvaliteten af genotypebestemmelsen (GQ) er under et skæringspunkt.</li> <li>• <b>q30</b> – Kvalitetsscore &lt; 30.</li> <li>• <b>LowVariantFreq</b> – Variantfrekvensen er mindre end den givne grænse.</li> <li>• <b>PB</b> – Probe pool bias. Variant ikke fundet eller fundet med lav frekvens i en eller to probepools.</li> <li>• <b>R3x6</b> – Antallet af tilstødende gentagelser (af en længde på 1 til 3 bp) til variantbestemmelserne <math>\geq 6</math>.</li> <li>• <b>SB</b> – Strand bias er over den givne grænse.</li> </ul>
<b>INFO</b>	<p>Mulige indtastninger i INFO-kolonnen omfatter:</p> <ul style="list-style-type: none"> <li>• <b>AC</b> – Alleltælling i genotyper for hver ALT-allel, i samme rækkefølge som anført.</li> <li>• <b>AF</b> – Allelfrekvens for hver ALT-allel, i samme rækkefølge som anført.</li> <li>• <b>AN</b> – Det samlede antal alleler i genotyper, der er bestemt.</li> <li>• <b>CD</b> – Et flag, der angiver, at SNP forekommer i koderegionen af mindst 1 RefGene-indtastning.</li> <li>• <b>DP</b> – Dybden (antal basebestemmelser alignet til en position og bruge i variantbestemmelse).</li> <li>• <b>Exon</b> – En kommasepareret liste over exonregioner, aflæst fra RefGene.</li> <li>• <b>FC</b> – Funktionel konsekvens.</li> <li>• <b>GI</b> – En kommasepareret liste over gen-ID'er, aflæst fra RefGene.</li> <li>• <b>QD</b> – Variantkonfidens/Dybdekvalitet.</li> <li>• <b>TI</b> – En kommasepareret liste over transkript-ID'er, aflæst fra RefGene.</li> </ul>
<b>FORMAT</b>	<p>Formatkolonnen angiver listefelter separeret af koloner. Eksempel: GT;GQ. Listen over felter afhænger af, hvilken Variant Caller der er brugt. Mulige tilgængelige felter:</p> <ul style="list-style-type: none"> <li>• <b>AD</b> – Indtastning af formelen X,Y, hvor X er antallet af referencebestemmelser, og Y er antallet af alternative bestemmelser.</li> <li>• <b>DP</b> – Omtrentlig læsningsdybde; læsninger med MQ=255 eller med "bad mates" filtreres ud.</li> <li>• <b>GQ</b> – Genotypekvalitet.</li> <li>• <b>GQX</b> – Genotypekvalitet. GQX er minimum af GQ-værdien og QUAL-kolonnen. Disse værdier er generelt identiske. Brug af minimum gør GQX til et mere konservativt mål for genotypekvaliteten.</li> <li>• <b>GT</b> – Genotype. 0 svarer til referencebasen, 1 svarer til den første indtastning i ALT-kolonnen og så videre. Skråstregen (/) angiver, at der ikke er tilgængelige fasningsoplysninger.</li> <li>• <b>NL</b> – Støjniveau; et estimate over basebestemmelsesstøjen på denne position.</li> <li>• <b>PB</b> – Probe pool bias. Værdier, der er nærmere 0, angav mere bias mod én probepool og mindre konfidens i en variantbestemmelse.</li> <li>• <b>SB</b> – Strand bias på denne position. Større negative værdier angiver mindre bias; værdier tæt på 0 angiver mere bias.</li> <li>• <b>VF</b> – Variantfrekvens; procentdelen af læsninger, der understøtter den alternative allel.</li> </ul>
<b>PRØVE</b>	Prøvekolonnen angiver de værdier, der er angivet i FORMAT-kolonnen.

## Genom-VCF-filer

Genom-VCF-filer (gVCF) er VCF v4.1-filer, der følger et sæt konventioner for repræsentation af alle sites i genomet i et forholdsvist kompakt format. gVCF-filen (\*.genom.vcf.gz) indeholder alle sites i det interessante område i en enkelt fil for hver prøve.

gVCF-filen viser ingen bestemmelser på positioner, der ikke passerer alle filtre. En genotype-tag (GT) på ./ angiver, at der ikke er sket nogen bestemmelse.

Læs mere på [sites.google.com/site/gvcftools/home/about-gvcf](https://sites.google.com/site/gvcftools/home/about-gvcf).

## Per-Pool- og konsensus-VCF-filer

Somatic Variant-arbejdsprocessen genererer 2 sæt variantbestemmelses- filer.

- ▶ **Per-pool VCF-filer** – Indeholder varianter, der er bestemt i enten forward pool eller reverse pool. Per-pool-filer skrives i VariantCallingLogs-mappen.
- ▶ **Konsensus VCF-filer** – Indeholder varianter, der er bestemt fra begge pools. Konsensus-filer skrives i Alignment-mappen.

Per-pool- og konsensus VCF-filer omfatter både VCF- (\*.vcf) og gVCF- (\*.genom.vcf) filer og bruger følgende navngivningskonvention, hvor S# repræsenterer den rækkefølge, hvori prøven er angivet til kørslen:

- ▶ **Rapporterer for alle sites** – Prøvenavn\_S#.genom.vcf
- ▶ **Rapporterer kun varianter** – Prøvenavn\_S#.vcf

Softwareen sammenligner per-pool VCF-filer og kombinerer dataene ved hver position for at skabe en konsensus VCF-fil for prøven.

Variantbestemmelser fra hver pool slås sammen i konsensus VCF-filer ud fra følgende kriterier.

Kriterier	Resultat
En referencebestemmelse i hver pool	Referencebestemmelse
En referencebestemmelse i 1 pool og en variantbestemmelse i den anden pool	Filtreret variantbestemmelse
Matchende variantbestemmelser med identiske frekvenser i hver pool	Variantbestemmelse
Matchende variantbestemmelser med signifikant forskellige frekvenser i hver pool	Filtreret variantbestemmelse
Ikke-matchede variantbestemmelser i hver pool	Filtreret variantbestemmelse

Måling fra hver pool slås sammen ud fra følgende værdier.

Måling	Værdi
Dybde	Tilføjelse af dybder fra begge pools
Variantfrekvens	Samlet antal varianter divideret med samlet dækningsdybde
Q-score	Minimumsværdi for begge pools

## Amplicondækningsfil

Der genereres en amplicondækningsfil til hver manifestfil. M# i filnavnet repræsenterer manifestnummeret. Hver fil har en toptekstrække med de prøve-ID'er, der er forbundet med manifestet. Filen indeholder følgende oplysninger.

- ▶ Mål-ID som anført i manifestet.
- ▶ Dækningsdybden af læsninger, der passerer filteret.

## Supplerende outputfiler

Følgende outputfiler giver supplerende oplysninger eller opsummerer kørselsresultater og analysefejl. Selvom disse filer ikke er nødvendige for vurdering af analyseresultater, kan de bruges til fejlfinding. Alle filer er placeret i Alignment-mappen, medmindre andet er angivet.

Filnavn	Beskrivelse
<b>AnalysisLog.txt</b>	Behandlingslog, der beskriver, hvert trin der forekom under analysen af den aktuelle kørselsmappe. Denne fil indeholder ikke fejlmeddelelser. Placeret i Alignmentmappen.
<b>AnalysisError.txt</b>	Behandlingslog, der anfører eventuelle fejl, der forekom under analysen. Denne fil vil være tom, hvis der ikke er sket fejl. Placeret i Alignmentmappen.
<b>DemultiplexSummaryF1L1#.txt</b>	Rapporterer demultiplexeringsresultater i en tabel med 1 række pr. flise og 1 kolonne pr. prøve. Nummeret repræsenterer bane 1, 2, 3 eller 4 i flowcellen. Placeret i Alignmentmappen.
<b>AmpliconRunStatistics.xml</b>	Indeholder oversigtsstatistik, der er specifik for kørslen. Placeret i Alignmentmappen.

## Analysemappen

Analysemappen indeholder de filer, der er genereret af Local Run Manager-softwaren.

Forholdet mellem udgangsmappen og analysemappen opsummeres som følger:

- ▶ Under sekventering udfylder Real-Time Analysis (RTA) udgangsmappen med filer, der er genereret under billedanalyse, basebestemmelse og kvalitetsscoreing.
- ▶ RTA kopierer filer til analysemappen i realtid. Når RTA har tildelt hver base en kvalitetsscore for hver cyklus, skriver softwaren filen RTAComplete.txt til begge mapper.
- ▶ Når filen RTAComplete.txt er til stede, begynder analysen.
- ▶ Idet analysen fortsætter, skriver Local Run Manager udgangsfilen til analysefilen og kopierer derefter filerne tilbage til udgangsfolderen.

## Alignmentmapper

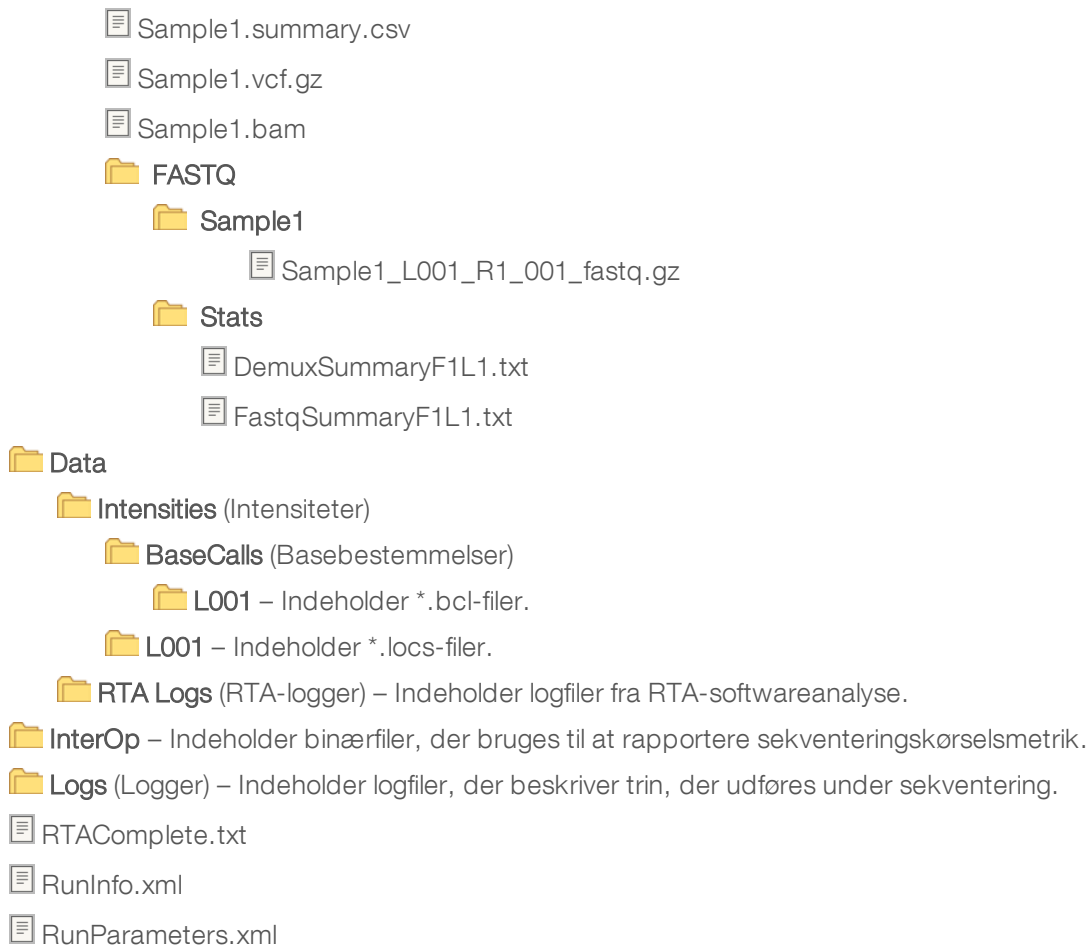
Hver gang denne analyse sættes i kø igen, opretter Local Run Manager en Alignmentmappe med navnet **Alignment\_N**, hvor N er et fortløbende tal.

## Mappestruktur

 **Alignment** – Indeholder filtyperne \*.bam, \*.vcf og FASTQ samt filer, der er specifikke for analysemodulet.

 **Dato- og tidsstempel** – Dato\_tidsstempel for analysen i formen ÅÅÅÅMMDD\_TTMMSS

-  AnalysisError.txt
-  AnalysisLog.txt
-  aggregate.report.html
-  aggregate.report.pdf
-  aggregate.summary.csv
-  AmpliconCoverage\_M#.tsv
-  AmpliconRunStatistics.xml
-  Sample1.genome.vcf.gz
-  Sample1.coverage.csv
-  Sample1.report.pdf



## Basebestemmelse og indeksdiversitet

Når prøver sekventeres på NextSeq 550Dx-instrumentet, bestemmer basebestemmelse en base (A, C, G eller T) for hver klynge af en given flise – billeddannelsesområde på flowcellen – i en given cyklus. NextSeq 550Dx-instrumentet bruger tokenalssekventering, der kun kræver to billeder til at indkode dataene til fire DNA-baser, én fra den røde kanal og én fra den grønne kanal.

Processen for basebestemmende indekxlæsninger er anderledes end basebestemmelse under andre læsninger.

Indekxlæsninger skal begynde med mindst én anden base end G i en af de første to cyklusser. Hvis en indekxlæsning begynder med to basebestemmelser af G, genereres der ingen signalstyrke. Der skal være signal i en af de første to cyklusser for at sikre demultiplekseringens ydelse.

Når der vælges indekser under oprettelse af kørslen, vises en advarsel om lav diversitet, hvis indekserne ikke opfylder kravene om diversitet. For at forhindre advarslen om lav diversitet skal der vælges indeksssekvenser, der giver signal i begge signaler for hver cyklus.

- ▶ Rød kanal – A eller C
- ▶ Grøn kanal – A eller T

Denne basebestemmelsesproces sikrer nøjagtighed ved analyse af lavpleks-prøver. Vedrørende yderligere oplysninger om sekvenserne af dine indekser henvises til indlægssedlen *TruSeq Custom Amplicon Kit Dx* (dokumentnr. 1000000029772).

Under oprettelse af kørsel i Local Run Manager skal du vælge det antal prøver, der skal testes. Foreslåede indeksskombinationer, der opfylder kravene til indeksdiversitet, bliver automatisk udfyldt af softwaren. Selvom du ikke skal bruge de foreslåede indeksskombinationer, anbefales det.



## Revisionshistorik

Dokument	Dato	Beskrivelse af ændring
Dokumentnr. 1000000030330 v04	August 2021	Opdateret adresse for EU-godkendt repræsentant.
Dokumentnr. 1000000030330 v03	April 2020	Opdateret adresse for EU-godkendt repræsentant. Opdateret adresse for australsk sponsor.
Dokumentnr. 1000000030330 v02	Januar 2019	Tilføjelse af oplysninger om v2.5-reagenskits.
Dokumentnr. 1000000030330 v01	August 2018	Opdatering af lovmæssige markeringer.
Dokumentnr. 1000000030330 v00	November 2017	Oprindelig udgivelse.

## Teknisk hjælp

Kontakt Illuminas tekniske 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 Illuminas kundesupport

Område	Gratis	Regional
Nordamerika	+1.800.809.4566	
Australien	+1.800.775.688	
Belgien	+32 80077160	+32 34002973
Danmark	+45 80820183	+45 89871156
Finland	+358 800918363	+358 974790110
Frankrig	+33 805102193	+33 170770446
Holland	+31 8000222493	+31 207132960
Hongkong	800960230	
Irland	+353 1800936608	+353 016950506
Italien	+39 800985513	+39 236003759
Japan	0800.111.5011	
Kina	400.066.5835	
New Zealand	0800.451.650	
Norge	+47 800 16836	+47 21939693
Østrig	+43 800006249	+43 19286540
Schweiz	+41 565800000	+41 800200442
Singapore	+1.800.579.2745	
Spanien	+34 911899417	+34 800300143
Storbritannien	+44 8000126019	+44 2073057197
Sverige	+46 850619671	+46 200883979
Taiwan	00806651752	
Tyskland	+49 8001014940	+49 8938035677
Andre lande	+44.1799.534000	

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

Produktokumentation – Kan downloades i PDF-format på Illuminas websted. Gå ind på [support.illumina.com](http://support.illumina.com), vælg et produkt, og vælg **Documentation & Literature**.



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

**KUN TIL IN VITRO-DIAGNOSTIK**

© 2021 Illumina, Inc. Alle rettigheder forbeholdes.

**illumina®**