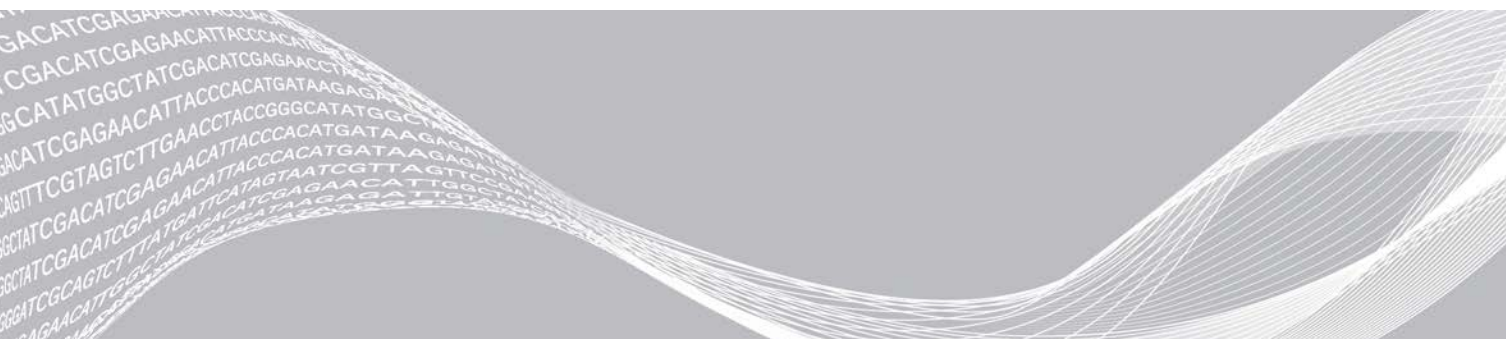


Local Run Manager Somatic Variant-analysemodul

Arbeidsflytveiledning for NextSeq 550Dx

TIL IN VITRO-DIAGNOSTISK BRUK

Oversikt	3
Oppgi kjøringsinformasjon	4
Analysemetoder	6
Vise kjørings- og prøvedata	7
Analyserapport	7
Analyseutdatafiler	9
Basebetegnelse og indeksmangfold	15
Revisjonslogg	17
Teknisk hjelp	18



Dette dokumentet og dets innhold er opphavsrettslig beskyttet for Illumina, Inc. og 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 strengt og tydelig 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).

© 2021 Illumina, Inc. Med enerett.

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

Oversikt

Local Run Manager Somatic Variant-modulen er tiltenkt bruk med Illumina TruSeq Custom Amplicon Kit Dx-analysen og NextSeq 550Dx. Når analysen brukes med Somatic Variant-modulen, er den ment for klargjøring av biblioteker som anvendes til sekvensering av DNA fra formalinfiksert, parafininnstøpt (FFPE) vev. Analysen oppdager somatiske mutasjoner ved lave variantfrekvenser.

Analysemodulen evaluerer korte regioner av amplifisert DNA, eller PCR-produkter, for varianter. Fokuset på sekvensering av amplikoner muliggjør høy dekning av bestemte regioner på tvers av et stort antall prøver. Analysemodulen utfører sekundær analyse og rapportgenerering fra sekvenseringskjøringer ved bruk av en to-strengt tilnærming som involverer en forover og en reversert oligo-pool. Se pakningsvedlegget *TruSeq Custom Amplicon Kit Dx* (dokumentnr. 1000000029772).

Somatic Variant-analysemodulen krever forbruksvarer for sekvensering med 300 sykluser. Du finner mer informasjon i pakningsvedleggene *NextSeq 550Dx High Output Reagent Kit v2* eller *NextSeq 550Dx High Output Reagent Kit v2.5*.

Om denne veiledningen

Denne veiledningen gir instruksjoner om å konfigurere kjøringsparametere for sekvensering og analyse for somatisk variantanalysemodulen. Informasjon om Local Run Manager-instrumentbordet og -systeminnstillingene finnes i *Referanseveiledning for NextSeq 550Dx-instrumentet* (dokumentnr. 1000000009513).

Vise Local Run Manager

Local Run Manager-grensesnittet vises i NextSeq 550Dx Operating Software (NOS) eller i en nettleser. Den støttede nettleseren er Chromium.



MERK

Hvis du bruker en nettleser som ikke støttes, må du laste ned den støttede nettleseren når du blir bedt om dette med meldingen «Confirm Unsupported Browser» (Bekreft ikke-støttet nettleser). Velg «**here**» (her) for å laste ned den støttede versjonen av Chromium.

Visning på instrumentmonitoren

- 1 Velg et av de følgende alternativene for å vise Local Run Manager-grensesnittet på instrumentmonitoren:
 - Fra startskjermbildet til NOS velger du **Local Run Manager**.
Klikk på symbolet X øverst i høyre hjørne for å gå tilbake til NOS når du er ferdig.
 - Velg ikonet for minimering av NOS, åpne Chromium-nettleseren på instrumentet, og tast deretter inn **http://localhost** i adresselinjen.
Kun brukere med administratorrettigheter kan minimere NOS.

Visning fra en datamaskin i et nettverk

- 1 Åpne en Chromium-nettleser på en datamaskin med tilgang til samme nettverk som instrumentet, og koble til ved hjelp av instrumentets IP-adresse eller instrumentnavn. For eksempel **http://myinstrument**.

Oppgi kjøringsinformasjon

Angi parametere

- 1 Logg deg på Local Run Manager.
- 2 Velg **Create Run** (Opprett kjøring), og velg **Somatic Variant**.
- 3 Skriv inn et kjøringsnavn som identifiserer kjøringen fra sekvensering gjennom analyse. Bruk alfanumeriske tegn, mellomrom, understreker eller bindestreker.
- 4 **[Valgfritt]** Angi en kjøringsbeskrivelse for å identifisere kjøringen. Bruk alfanumeriske tegn, mellomrom, understreker eller bindestreker.
- 5 Velg antall prøver og indekssett fra rullegardinlisten. Ta hensyn til følgende informasjon når du foretar et valg.
 - ▶ Rullegardinlisten inneholder antall prøver med et indekssett. For eksempel indikerer «24-Set 1» at 24 prøver skal testes, med indekser fra indekssett 1.
 - ▶ Indekssettnumre refererer til ulike sett med i5-indekser. Både sett 1 og sett 2 gir indeksangfold. To indekssett tilbys for å bidra til å forhindre uttømming av et enkelt sett.
 - ▶ Velg antall prøver som ligger nærmest antall prøver som du skal teste. Hvis nøyaktig antall prøver ikke finnes i listen, velger du det nærmeste antallet, men mindre enn antallet du skal teste. Hvis du for eksempel ønsker å teste 18 prøver, velger du 16 prøver.
 - ▶ Prøvebrønner og indekskombinasjoner som oppfyller indeksangfoldkrav, utheves med grønt. Hvis du velger andre brønner og indekskombinasjoner, vises det en melding når du lagrer kjøringen, dersom indeksangfoldkravene ikke er oppfylt.

Importere manifestfiler for kjøringen

- 1 Kontroller at manifestene du vil importere, er tilgjengelige på en tilgjengelig nettverksplasing eller på en USB-stasjon.
- 2 Velg **Import Manifests** (Importer manifeste).
- 3 Naviger til manifestfilen, og velg manifestene du vil legge til.



MERK

For å gjøre manifestfiler tilgjengelige for alle kjøringer ved hjelp av den somatiske analysemodulen legger du til manifestene ved hjelp av modulinnstillingsfunksjonen. Denne funksjonen krever administratorrettigheter. Du finner mer informasjon i *Referanseveiledning for NextSeq 550Dx-instrumentet (dokumentnr. 1000000009513)*.

Angi prøver for kjøringen

Angi prøver for kjøringen ved å bruke ett av alternativene og instruksjonene som følger.

- ▶ **Enter samples manually** (Angi prøver manuelt) – Bruk den tomme tabellen i skjermbildet Create Run (Opprett kjøring).
- ▶ **Import samples** (Importer prøver) – Naviger til en ekstern fil i et format med kommaseparerte verdier (*.csv). En mal er tilgjengelig for nedlasting i skjermbildet Create Run (Opprett kjøring).

Etter at du har fylt ut prøvetabellen, kan du eksportere prøveinformasjonen til en ekstern fil. Bruk filen som referanse når du klargjør biblioteker eller importerer filen for en ny kjøring.

Legge inn prøver manuelt

- 1 Angi et unikt prøvenavn i feltet Sample Name (Prøvenavn).
Bruk alfanumeriske tegn, bindestreker eller understreker.
Prøvenavnet fyller automatisk den tilsvarende brønnen i den andre poolen.
- 2 **[Valgfritt]** For positive eller negative kontrollprøver, høyreklikk og velg kontrolltypen.
Kontrollen i en prøvebrønn fyller automatisk den tilsvarende brønnen i den andre poolen med samme kontroll.
- 3 **[Valgfritt]** Skriv inn en prøvebeskrivelse i feltet Sample Description (Prøvebeskrivelse).
Bruk alfanumeriske tegn, bindestreker eller understreker.
Prøvebeskrivelsen fyller automatisk den tilsvarende brønnen i den andre poolen.
Prøvebeskrivelser er knyttet til en prøve-ID. Prøvebeskrivelser overskrives hvis samme prøve-ID brukes igjen i en senere kjøring.
- 4 Velg en indeks 1-adapter fra rullegardinlisten Index 1 (i7) (Indeks 1 (i7)).
Når du bruker foreslåtte prøvebrønner, fyller programvaren automatisk ut i7- og i5-indeksadaptere som oppfyller mangfoldindekskrav. Hvis det nøyaktige antallet prøver du tester, ikke er på listen, må du påse å velge indeksadaptere for ekstra brønner. Dersom du må velge indekser for ekstra brønner, eller dersom du ikke bruker de anbefalte indeksadapterkombinasjonene, må du sørge for å lese avsnittet *Basebetegnelse og indeksmangfold på side 15* før du velger indekser.
- 5 Velg en indeks 2-adapter fra rullegardinlisten Index 2 (i5) (Indeks 2 (i5)).
- 6 Velg en manifestfil fra rullegardinlisten Manifest.
Prøver i pool A krever et annet manifest enn prøver i pool B.
- 7 Velg et alternativ for å vise, skrive ut eller lagre plateoppsettet som en referanse for klargjøring av biblioteker:
 - ▶ Velg ikonet  **Print** (Skriv ut) for å vise plateoppsettet. Velg **Print** (Skriv ut) for å skrive ut plateoppsettet.
 - ▶ Velg **Export** (Eksporter) for å eksportere prøveinformasjon til en ekstern fil.
Kontroller at manifest- og prøveinformasjon er riktig. Feil informasjon kan påvirke resultatene.
- 8 Velg **Save Run** (Lagre kjøring).

Importere prøver

- 1 Velg **Import Samples** (Importer prøver), og bla til plasseringen av prøveinformasjonsfilen. Det finnes to typer filer du kan importere.
 - ▶ Velg **Template** (Mal) i skjermbildet Create Run (Opprett kjøring) for å lage et nytt plateoppsett.
Malfilen inneholder de riktige kolonneoverskriftene for import. Oppgi prøveinformasjon i hver kolonne for prøvene i kjøringen. Slett eksempelinformasjon i ubrukte celler, og lagre deretter filen.
 - ▶ Bruk en fil med prøveinformasjon som ble eksportert fra Somatic Variant-modulen med funksjonen Export (Eksporter).
- 2 Velg ikonet  **Print** (Skriv ut) for å vise plateoppsettet.
- 3 Velg **Print** (Skriv ut) for å skrive ut plateoppsettet som en referanse for klargjøring av biblioteker.
- 4 **[Valgfritt]** Velg **Export** (Eksporter) for å eksportere prøveinformasjon til en ekstern fil.

Kontroller at manifest- og prøveinformasjon er riktig. Feil informasjon kan påvirke resultatene.

5 Velg **Save Run** (Lagre kjøring).

Redigere en kjøring

Instruksjoner om hvordan du redigerer informasjonen i kjøringen før sekvensering, finnes i *Referanseveiledning for NextSeq 550Dx-instrumentet* (dokumentnr. 1000000009513).

Analysemetoder

Den somatiske variantanalysemodulen utfører følgende analysetrinn og skriver deretter analyseutdatafiler til mappen Alignment (Innretting).

- ▶ Demultiplekser indeksavlesninger
- ▶ Genererer FASTQ-filer
- ▶ Innretter etter en referanse
- ▶ Identifiserer varianter

Demultipleksing

Demultipleksing sammenligner hver indeksavlesningssekvens med indekssekvensene som er angitt for kjøringen. Ingen kvalitetsverdier vurderes på dette trinnet.

Indeksavlesninger identifiseres ved hjelp av følgende trinn:

- ▶ Prøvene er nummerert fra 1 basert på rekkefølgen de er oppført i for kjøringen.
- ▶ Prøvenummer 0 er reservert for klynger som ikke ble tildelt en prøve.
- ▶ Klynger tilordnes en prøve når indekssekvensen samsvarer nøyaktig eller når det er opptil ett misforhold per indeksavlesning.

FASTQ-filgenerering

Etter demultipleksing genererer programvaren mellomliggende analysefiler i FASTQ-formatet, som er et tekstformat som brukes til å representere sekvenser. FASTQ-filer inneholder avlesninger for hver prøve og tilhørende kvalitetsscore. Klynger som ikke passerte filter, utelukkes.

Hver FASTQ-fil inneholder avlesninger for bare én prøve, og navnet på denne prøven er inkludert i FASTQ-filnavnet. FASTQ-filer er primære inndata for innretting. Åtte FASTQ-filer genereres per prøve per oligo-pool, fire fra Read 1 og fire fra Read 2, noe som resulterer i 16 FASTQ-filer per prøve.

Innretting

Under innrettingstrinnet innretter den bundne Smith-Waterman-algoritmen klynger fra hver prøve med PCR-produktsekvenser spesifisert i manifestfilen.

Den bundne Smith-Waterman-algoritmen utfører halvglobal sekvensinnretting for å bestemme lignende regioner mellom to sekvenser. I stedet for å sammenligne den totale sekvensen, sammenligner Smith-Waterman-algoritmen segmenter med alle mulige lengder.

Hver paired-end-avlesning blir evaluert med hensyn til sin innretting med de relevante probesekvensene for denne avlesningen.

- ▶ Read 1 evalueres mot reverskomplementet av nedstrøms lokusspesifikke oligoer (DLSO).
- ▶ Read 2 blir vurdert opp mot oppstrøms lokusspesifikke oligoer (ULSO).

- ▶ Hvis starten på en avlesing samsvarer med en probesekvens med ikke mer enn tre ulikheter (misforhold eller forskyvninger på grunn av ledende indeler), innrettes den fulle lengden av avlesingen mot PCR-produktmålet for den sekvensen.
- ▶ Indeler innenfor DLSO og ULSO blir ikke observert på grunn av analysekjemien.

Innrettinger filtreres fra innrettingsresultater basert på antall misforhold over enten interesseområdet eller hele PCR-produktet, avhengig av PCR-produktlengden. Filtrerte innrettinger registreres i innrettingsfiler som uinnrettede og brukes ikke i variantbetegnelse.

Variantbetegnelse

Pisces variantbetegner er utviklet av Illumina og identifiserer varianter som er til stede med lav frekvens i DNA-prøven.

Pisces variantbetegner identifiserer SNV-er, MNV-er og små indeler i tre trinn:

- ▶ Vurderer hver posisjon i referanseggenomet separat
- ▶ Teller baser på den angitte posisjonen for innrettede avlesninger som overlapper posisjonen
- ▶ Beregner en variantscore som måler kvaliteten på betegnelsen ved hjelp av Poisson-modellen. Varianter med kvalitetsscore under Q30 utelates.

Varianter betegnes først for hver pool hver for seg. Deretter sammenlignes og kombineres varianter fra hver pool til en enkelt utdatafil. Hvis en variant er til stede i begge poolene og passerer alle filtrene oppført i *VCF-filkommentarer på side 12*, merkes varianten som PASS i variantbetegnelsesfilen (VCF).

Vise kjørings- og prøvedata

- 1 Fra Local Run Manager-instrumentbordet klikker du på kjøringsnavnet.
- 2 Fra fanen Run Overview (Kjøringsoversikt) går du gjennom sekvenskjøringsmetrikken.
- 3 **[Valgfritt]** Klikk på ikonet **Copy to Clipboard**  (Kopier til utklippstavle) for å kopiere banen til utgangskjøringsmappen.
- 4 Klikk på fanen Sequencing Information (Sekvenseringsinformasjon) for å se informasjon om kjøringsparametere og forbruksvarer.
- 5 Klikk på fanen Samples and Results (Prøver og resultater) for å vise analyserapportens plassering.
 - ▶ Hvis analysen ble gjentatt, utvider du rullegardinlisten Select Analysis (Velg analyse) og velger den aktuelle analysen.
- 6 Klikk på ikonet **Copy to Clipboard**  (Kopier til utklippstavle) for å kopiere banen til mappen Analysis (Analyse).

Du finner mer informasjon om fanene Run Overview (Kjøringsoversikt) og Sequencing Information (Sekvenseringsinformasjon) og om hvordan analyser settes tilbake i kø, i *Referanseveiledning for NextSeq 550Dx-instrumentet (dokumentnr. 1000000009513)*.

Analyserapport

Analyseresultater oppsummeres på fanen Samples and Results (Prøver og resultater) og som en samlet rapport i mappen Alignment (Innretting). En rapport for hver prøve er også tilgjengelig i PDF-filformat for hver prøve.

Informasjon om fanen Samples and Results (Prøver og resultater)

- 1 Klikk på en prøve i listen for å se prøverapporten.

Tabell 1 Kjørings- og prøveinformasjon

Kolonneoverskrift	Beskrivelse
Run Status (Kjøringsstatus)	Angir om sekvenseringskjøringen var godkjent eller mislyktes.
Total Yield (GB) (Total produksjon (GB))	Antall baser betegnet i sekvenseringskjøringen. Viser grenseverdi for godkjenning, og status (godkjent eller mislykket).
% ≥ Q30	Prosentandelen av avlesninger i sekvenseringskjøringen med en kvalitetsscore på 30 (Q30) eller høyere. Viser grenseverdi for godkjenning, og status (godkjent eller mislykket).
Sample Name (Prøvenavn)	Prøvenavnet som ble gitt da kjøringen ble opprettet.
Total PF Reads (PF-avlesninger totalt)	Totalt antall avlesninger som passerer filteret.
Read 1% ≥ Q30	Prosentandelen av avlesninger i Read 1 med en kvalitetsscore på 30 (Q30) eller høyere for prøven.
Read 2% ≥ Q30	Prosentandelen av avlesninger i Read 2 med en kvalitetsscore på 30 (Q30) eller høyere for prøven.
Autosome Call Rate (Betegnelsesfrekvens for autosomer)	Antallet genomiske posisjoner blant autosomene (kromosom 1 til 22) som oppfyller en forhåndsdefinert konfidensverditerkel, dividert med det totale antallet autosomale genomiske posisjoner som behandles. Betegnelsesfrekvensen på per prøve-basis og rapportert som prosentandel som er beregnet som 1 minus (antall autosomale posisjoner med ufullstendige betegnelser, dividert med totalt antall sekvenserte autosomale posisjoner).

Tabell 2 Informasjon om prøverapport

Kolonneoverskrift	Beskrivelse
Sample (Prøve)	Prøvenavnet som ble gitt da kjøringen ble opprettet.
Report Date (Rapportdato)	Datoen rapporten ble generert.
Sample Information (Prøveinformasjon)	Prøve-ID-en som ble oppgitt da kjøringen ble opprettet, totalt antall avlesninger som passerte filteret i prøven, prosentandelen av avlesninger for prøven med en kvalitetsscore på 30 (Q30) eller høyere, og den autosomale betegnelsesfrekvensen.
Amplicon Summary (PCR-produktsammendrag)	Totalt antall sekvenserte PCR-produktregioner, og total lengde i basepar av sekvenserte PCR-produkter i målregionene, for prøven i pool A og pool B og manifestfilen som brukes for hver pool. Manifestfilen angir referanseggenomet og målrettede referanseregioner som brukes i innrettingstrinnet.
Read Level Statistics (Statistikk på avlesningsnivå)	Antall og prosent av avlesninger for prøven som dekker hver posisjon i referansen, for Read 1 og Read 2 i pool A og pool B.
Variants Summary (Variantsammendrag)	Antall SNV-er, innsettinger og slettinger oppdaget for prøven som passerte foreslåtte verdier, for å avgjøre om kvalitetsresultatene er innenfor et akseptabelt område,
Coverage Summary (Dekningsammendrag)	Det totale antallet justerte baser dividert med målrettet regionstørrelse og prosentandelen av PCR-produktregioner med dekningsverdier som er større enn den lave dekningssterskelen på 0,2* PCR-produktdekning, for prøven i pool A og pool B.

Kolonneoverskrift	Beskrivelse
Coverage Plots (Dekningsplott)	Dekningen av PCR-produktregionplottet viser dekingen over PCR-produktregionene for prøven. Regioner med dekingverdier som er lavere enn dekingsterskelen, er markert i rødt. Gjennomsnittet av alle verdiene er angitt med en oransje linje. Et plott er gitt for deking av pool A og pool B.
Software Versions (Programvareversjoner)	Programvareversjoner da prøven ble sekvensert. Inkluderer NextSeq 550Dx Operating Software (NOS), Local Run Manager Software, RTA Software og Somatic Variant-modulversjonen.

Analyseutdatafiler

Følgende analyseutdatafiler genereres for den somatiske variantanalysemodulen og gir analyseresultater for innretting og variantbetegnelse. Analyseutdatafiler finnes i mappen Alignment (Innretting).

Filnavn	Beskrivelse
Demultipleksing (*.txt)	Mellomliggende filer som inneholder sammendrag av resultater fra demultipleksing.
FASTQ (*.fastq.gz)	Mellomliggende filer som inneholder kvalitetsscorede basebetegnelser. FASTQ-filer er den primære inngangen for innrettingstrinnet.
Innrettingsfiler i BAM-format (*.bam)	Inneholder innrettede avlesinger for en gitt prøve.
Filer for variantbetegnelse per pool i VCF-format (*.vcf)	Inneholder varianter som betegnes i hver posisjon fra enten forover pool eller reversert pool.
Variantbetegnelsesfiler i genom-VCF-format (*.genome.vcf.gz)	Inneholder genotypen for hver posisjon, enten den betegnes som en variant eller som en referanse.
Konsensus-variantbetegnelsesfiler i VCF-formatet (*.vcf.gz)	Inneholder varianter som betegnes ved hver posisjon fra begge pool.
AmpliconCoverage_M1.tsv	Inneholder informasjon om deking per PCR-produkt per prøve for hvert oppgitte manifest. M# angir manifesttallet.

Filformat for demultipleksing

Demultipleksingprosessen leser indekssekvensen for hver klynge for å bestemme hvilken prøve klyngen oppsto fra. Kartleggingen mellom klynger og prøvenummer skrives til en demultipleksingsfil (*.demux) for hver plate av strømningscellen.

Filnavnet for demultipleksingsfiler er **s_1_X.demux**, hvor X er platenummeret.

Demultipleksingsfiler starter med en overskrift:

- ▶ Versjon (4 byte heltall), for tiden 1
- ▶ Klyngetall (4 byte heltall)

Resten av filen består av prøvetall for hver klynge fra platen.

Når demultipleksingstrinnet er fullført, genererer programvaren en demultipleksingsfil som heter **DemultiplexSummaryF1L1.txt**.

- ▶ I filnavnet representerer **F1** strømningscellenummeret.
- ▶ I filnavnet representerer **L1** banenummeret.
- ▶ Demultipleksingsresultater i en tabell med én rad per plate og én kolonne per prøve, inkludert prøve 0.
- ▶ De vanligste sekvensene i indeksavlesninger.

FASTQ-filformat

FASTQ er et tekstbasert filformat som inneholder basebetegnelser og kvalitetsverdier per avlesning. Hver oppføring inneholder 4 linjer:

- ▶ Identifikatoren
- ▶ Sekvensen
- ▶ Et plusstegn (+)
- ▶ Phred-kvalitetsresultatene i et ASCII + 33-kodet format

Identifikatoren er formaterert som:

@Instrument:Kjørings-ID:Strømningscelle-ID:Bane:Plate:X:Y Avlesningsnr.:FilterFlagg:0:Prøvenummer

Eksempel:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=#
```

BAM-filformat

En BAM-fil (*.bam) er en komprimert binær versjon av en SAM-fil som brukes til å angi innrettede sekvenser opptil 128 Mb. SAM- og BAM-formater beskrives i detalj på samtools.github.io/hts-specs/SAMv1.pdf.

BAM-filer bruker filnavngivningsformatet **SampleName_S#.bam**, der # er prøvenummeret som bestemt av rekkefølgen av prøvene som er oppført for kjøringen.

BAM-filer inneholder et overskriftsavsnitt og et innrettingsavsnitt:

- ▶ **Header** (Overskrift) – Inneholder informasjon om hele filen, som prøvenavn, prøvelengde og innrettingsmetode. Innrettinger i innrettingsavsnittet er forbundet med spesifikk informasjon i overskriftsavsnittet.
- ▶ **Alignments** (Innrettinger) – Inneholder avlesningsnavn, avlesningssekvens, avlesningskvalitet, innrettingsinformasjon og tilpassede faner. Avlesningsnavnet inkluderer kromosomet, startkoordinat, innrettingskvalitet og samsvarsindikatorstreng.

Innrettingsavsnittet inneholder følgende informasjon for hver avlesning eller hvert avlesningspar:

- ▶ **AS:** Paired-end innrettingskvalitet
- ▶ **BC:** Strekkode-fane, som indikerer den demultipleksede prøve-ID-en som er knyttet til avlesningen.
- ▶ **SM:** Single-end innrettingskvalitet.
- ▶ **XC:** Samsvarsindikatorstreng
- ▶ **XN:** Navnefaner for PCR-produkt, som registrerer PCR-produkt-ID-en som er knyttet til avlesningen

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

VCF-filformat

Variant Call Format (VCF) er et vanlig filformat utviklet av det genomikkvitenskapelige samfunnet. Det inneholder informasjon om varianter funnet i spesifikke posisjoner i et referansegenom. VCF-filer slutter med .vcf-suffikset

VCF-filoverskriften inneholder VCF-filformatversjonen og variantbetegnersversjonen og angir kommentarene som brukes i resten av filen. VCF-overskriften inneholder også referansegenomfilen og BAM-filen. Den siste linjen i overskriften inneholder kolonneoverskriftene for datalinjene. Hver av VCF-fildatalinjene inneholder informasjon om én variant.

VCF-filoverskrifter

Overskrift	Beskrivelse
CHROM (KROM)	Kromosomet på referansegenomet. Kromosomer vises i samme rekkefølge som referanse-FASTQ-filen.
POS	Enkeltbaseposisjonen til varianten i referansechromosomet. I SNP-er er denne posisjonen referansebasen med varianten, i indeler eller slettinger er denne posisjonen referansebasen umiddelbart før varianten.
ID	rs-nummeret for varianten, hentet fra dbSNP.txt, hvis dette er aktuelt. Hvis det finnes flere rs-numre på denne plasseringen, blir listen avgrenset med semikolon. Hvis det ikke eksisterer noen dbSNP-oppføring i denne posisjonen, brukes det en markør for manglende verdi ('.').
REF	Referansegenotypen. Eksempel: En sletting av en enkel T angis som referanse-TT og alternativ T. En enkel A til T-nukleotidvariant angis som referanse-A og alternativ T.
ALT	Allelene som er forskjellige fra referanseavlesingen. Eksempel: en innsetting av en enkel T angis som referanse-A og alternativ AT. En enkel A til T-nukleotidvariant er representert som referanse-A og alternativ T.
QUAL (KVAL)	En Phred-skalert kvalitetsscore tildelt av variantbetegneren. Høyere score angir høyere konfidens i varianten og lavere sannsynlighet for feil. I en kvalitetsscore på Q er den anslåtte sannsynligheten for en feil $10^{-(Q/10)}$. Eksempel: settet med Q30-betegnelser har en feilrate på 0,1 %. Mange variantbetegnere tildeler kvalitetsscorer basert på de tilhørende statistikkmodellene, som er høye i forhold til den observerte feilraten.

VCF-filkommentarer

Overskrift	Beskrivelse
FILTER	Hvis alle filtrene godkjennes, blir PASS skrevet i filterkolonnen. • LowDP – Gjelder steder der dekningsdybden er under 450x i en av poolene. For PCR-produktposisjoner som dekkes av avlesning både forover og reversert, svarer dette til 900x enkeltlesningsdekning. • LowGQ – Genotypingkvaliteten (GQ) er under en cut-off. • q30 – Kvalitetsscore < 30. • LowVariantFreq – Variantfrekvensen er lavere enn den angitte terskelen. • PB – Probepool-avvik. Variant ikke funnet, eller funnet med lav frekvens i en eller to probepooler. • R3x6 – Antall tilstøtende repetisjoner (med lengde 1 til 3 bp) til variantbetegnelse ≥ 6. • SB – Strengeavviket er mer enn gitt grenseverdi.
INFO	Mulige oppføringer i INFO-kolonnen omfatter: • AC – Alleltelling i genotyper for hver ALT-allel, i samme rekkefølge som oppført. • AF – Allelfrekvens for hver ALT-allel, i samme rekkefølge som oppført. • AN – Det totale antallet alleler i betegnede genotyper. • CD – Et flagg som indikerer at SNP opptrer innenfor kodingsområdet med minst 1 RefGene-oppføring. • DP – Dybden (antall basebetegnelser som er innrettet til en posisjon og brukt i variantbetegnelse). • Exon – En kommaseparert liste over eksonregioner avlest fra RefGene. • FC – Funksjonell konsekvens. • GI – En kommaseparert liste over gen-ID-er avlest fra RefGene. • QD – Variantkonfidens / kvalitet etter dybde. • TI – En kommaseparert liste over transkripsjons-ID-er avlest fra RefGene.
FORMAT	Formatkolonnen viser felt adskilt med kolon. For eksempel GT:GQ. Listen over viste felter avhenger av variantbetegneren som ble brukt. Følgende felt er tilgjengelige: • AD – Oppføring i skjemaet X, Y, der X er antallet referansebetegnelser, og Y er antallet alternative betegnelser. • DP – Omtrentlig avlesingsdybde, avlesinger med MQ=255 eller med feil tilordninger blir filtrert. • GQ – Genotypekvalitet. • GQX – Genotypekvalitet. GQX er minimum av GQ-verdien og QUAL-kolonnen. Generelt sett er disse verdiene like. Bruk av minimum gjør GQX det mest forsiktede målet på genotypekvalitet. • GT – Genotype. 0 tilsvarer referansebasen, 1 tilsvarer den første oppføringen i ALT-kolonnen, osv. Skråstrek (/) angir at ingen fasingsinformasjon er tilgjengelig. • NL – Støynivå; et estimat av basebetegnelsestøy i denne posisjonen. • PB – Probepool-avvik. Verdier nærmere 0 indikerte større avvik mot én probepool, og mindre konfidens i en variantbetegnelse. • SB – Strengeavvik på denne posisjonen. Større negative verdier indikerer mindre avvik; verdier nær 0 angir mer avvik. • VF – Variantfrekvens, prosentandelen av avlesingene som støtter den alternative allelen.
SAMPLE (PRØVE)	I prøvekolonnen oppgis verdiene som angis i FORMAT-kolonnen.

Genome VCF-filer

Genome VCF (gVCF)-filer er VCF v4.1-filer som følger et sett med konvensjoner for å representere alle områder i genomet i et rimelig kompakt format. gVCF-filene (*.genome.vcf.gz) inneholder alle områdene i interesseregionen i en enkelt fil for hver prøve.

gVCF-filen viser ikke-påvisninger på posisjoner som ikke passerer alle filtre. En genotype (GT)-fane med ./ indikerer en ikke-påvisning.

Se sites.google.com/site/gvcftools/home/about-gvcf for mer informasjon.

VCF-filer per pool og konsensus

Somatisk variant-arbeidsflyten genererer to sett med variant betegnelsesfiler.

- ▶ **VCF-filer per pool** – Inneholder varianter betegnet i enten forover pool eller reversert pool. Filer per pool skrives til mappen VariantCallingLogs (Variantbetegnelseslogger).
- ▶ **Konsensus VCF-filer** – Inneholder varianter betegnet fra begge pooler. Konsensus-filer skrives til mappen Alignment (Innretting).

VCF-filer (VCF) per pool og konsensus omfatter både VCF (*.vcf) og gVCF (*.genome.vcf)-filer, og bruker følgende navngivningskonvensjon, hvor S# representerer rekkefølgen som prøven er oppført i for kjøringen:

- ▶ **Rapporterer for alle steder** – SampleName_S#.genome.vcf
- ▶ **Rapporterer kun varianter** –SampleName_S#.vcf

Programvaren sammenligner VCF-filene per pool og kombinerer dataene i hver posisjon for å lage en konsensus-VCF-fil for prøven.

Variantbetegnelser fra hver pool slås sammen til konsensus-VCF-filer ved hjelp av følgende kriterier.

Kriterier	Resultat
En referansebetegnelse i hver pool	Referansebetegnelse
En referansebetegnelse i en pool og en variantbetegnelse i den andre poolen	Filtrert variantbetegnelse
Matcher variantbetegnelser med lignende frekvenser i hver pool	Variantbetegnelse
Matcher variantbetegnelser med betydelig forskjellige frekvenser i hver pool	Filtrert variantbetegnelse
Umatchede variantbetegnelser i hver pool	Filtrert variantbetegnelse

Metrikk fra hver pool slås sammen med følgende verdier.

Metrikk	Verdi
Dybde	Tilsetning av dybder fra begge pooler
Varianthypighet	Totalt variantantall dividert med total dekningsdybde
Q-score	Minste verdi av begge pool

Fil for PCR-produktdekning

Én fil for PCR-produktdekning genereres for hver manifestfil. M# i filen angir manifesttallet.

Hver fil omfatter en overskriftsrad som inneholder prøve-ID-ene forbundet med manifestet. Filen inneholder følgende informasjon.

- ▶ Mål-ID-en slik den er oppført i manifestet.
- ▶ Dekningsdybden av avlesninger som passerer filteret.

Tilleggsutdatafiler

Følgende utdatafiler gir tilleggsinformasjon, eller oppsummerer kjøringresultater og analysefeil. Selv om disse filene ikke er påkrevd for å vurdere analyseresultater, kan de brukes for feilsøkningsformål. Alle filene ligger i mappen Alignment (Innretting) med mindre annet er angitt.

Filnavn	Beskrivelse
AnalysisLog.txt	Behandlingslogg som beskriver hvert trinn som oppstod under analysen av gjeldende kjørsmappe. Denne filen inneholder ingen feilmeldinger. Ligger i mappen Alignment (Innretting).
AnalysisError.txt	Behandlingsloggen som har opplistet eventuelle feil som har oppstått under analyse. Denne filen vil være tom hvis det ikke oppsto noen feil. Ligger i mappen Alignment (Innretting).
DemultiplexSummaryF1L1#.txt	Rapporterer demultipleksingsresultater i en tabell med 1 rad per plate og 1 kolonne per prøve. # representerer bane 1, 2, 3 eller 4 av strømningscellen. Ligger i mappen Alignment (Innretting).
AmpliconRunStatistics.xml	Inneholder oppsummeringsstatistikk som er spesifikk for kjøringen. Ligger i mappen Alignment (Innretting).

Analysemappe

Analysemappen inneholder filene som genereres av Local Run Manager-programvaren.

Forholdet mellom utgangsmappen og analysemappen er oppsummert som følger:

- ▶ Under sekvensering fyller Real-Time Analysis (RTA) utgangsmappen med filer som genereres under bildeanalyse, basebetegnelse og kvalitetsscoreing.
- ▶ RTA kopierer filer til analysemappen i sanntid. Etter at RTA har tilordnet en kvalitetsscore til hver base for hver syklus, skriver programvaren filen RTAComplete.txt til begge mappene.
- ▶ Når filen RTAComplete.txt er til stede, begynner analysen.
- ▶ Mens analysen fortsetter, skriver Local Run Manager utdatafiler til analysemappen og kopierer deretter filene til utgangsmappen.

Innrettingsmapper

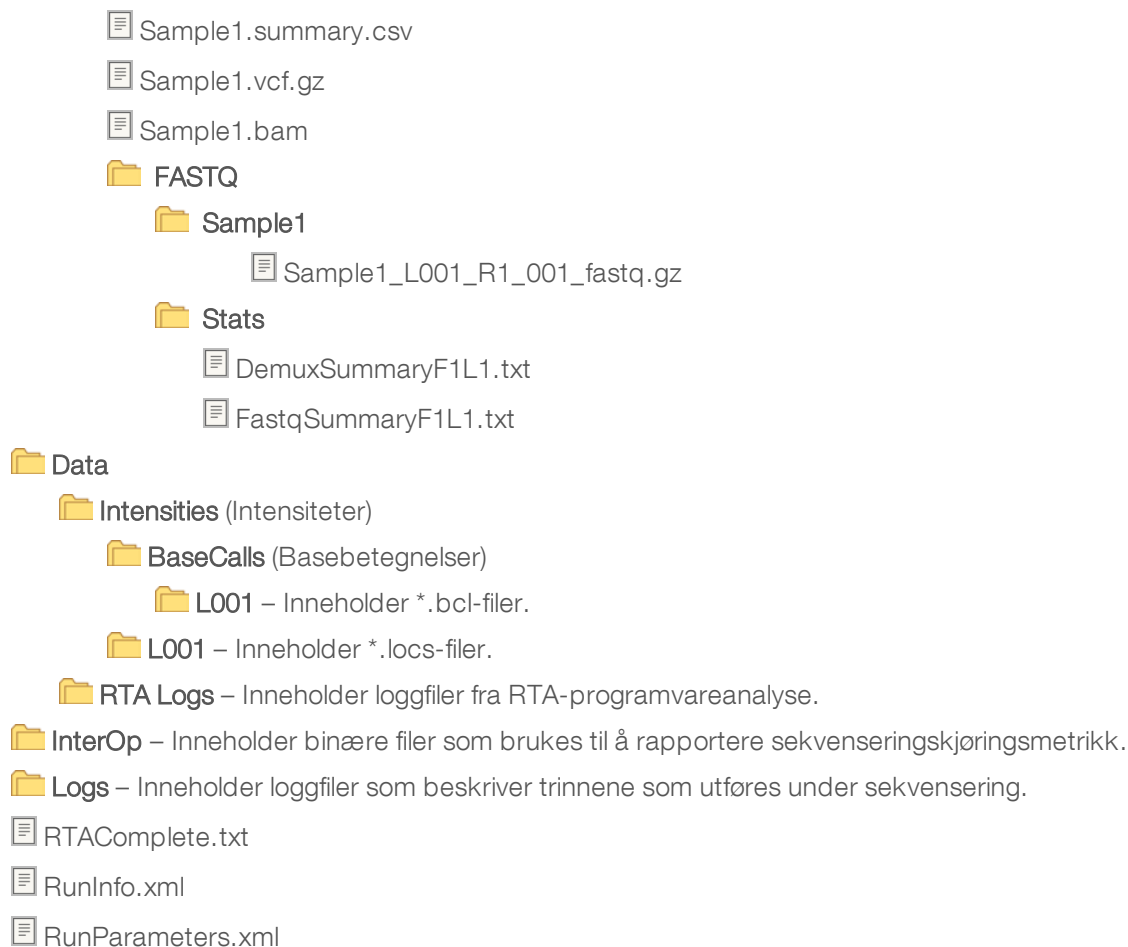
Hver gang denne analysen settes tilbake i kø, oppretter Local Run Manager en innrettingsmappe som heter **Alignment_N** (Innretting_N), hvor N er et sekvensielt tall.

Mappestruktur

 **Alignment** – Inneholder *.bam-, *.vcf-, FASTQ-filer, og filer som er spesifikke for analysemodulen.

 **Dato og klokkeslettstempel** – Date_time-stempel av analyse som 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



Basebetegnelse og indeksmangfold

Når prøver sekvenseres med NextSeq 550Dx-instrumentet, vil basebetegnelsen bestemme en base (A, C, G eller T) for hver klynge til en bestemt plate, eller avbildningsområdet på strømningscellen, ved en spesifikk syklus. NextSeq 550Dx-instrumentet bruker tokenalssekvensering som kun krever to bilder for å kode dataene for fire DNA-baser: ett fra den røde kanalen og ett fra den grønne kanalen.

Proessen for basebetegnelsesindeksavlesninger avviker fra basebetegnelse ved andre avlesninger.

Indeksavlesninger må starte med minst én base som ikke er G, i én av de to første syklusene. Hvis en indeksavlesning starter med to G-basebetegnelser, genereres det ingen signalintensitet. Signalet må være til stede i én av de to første syklusene for å kunne sikre demultipleksing.

Når du velger indekser ved oppretting av en kjøring, vil det vises en advarsel om lavt mangfold dersom indeksene ikke oppfyller mangfoldkravene. For å unngå en advarsel om lavt mangfold må du velge indeksssekvenser som gir signal i begge kanalene for hver syklus.

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

Denne basebetegnelsesprosessen sikrer nøyaktighet ved analysing av få prøver samtidig. Du finner mer informasjon om sekvensering av indekser i pakningsvedlegget *TruSeq Custom Amplicon Kit Dx (dokumentnr. 1000000029772)*.

Når du oppretter en kjøring i Local Run Manager, må du velge antallet prøver som skal testes. Foreslåtte indeksskombinasjoner som oppfyller indeksmangfoldkravene, fylles automatisk ut av programvaren. Selv om det ikke er påkrevd å bruke foreslåtte indeksskombinasjoner, anbefales det.

Revisjonslogg

Dokument	Dato	Beskrivelse av endring
Dokumentnr. 1000000030330 v04	August 2021	Oppdatert adresse for EU-autorisert representant.
Dokumentnr. 1000000030330 v03	April 2020	Oppdatert adresse for EU-autorisert representant. Oppdatert adresse for australsk sponsor.
Dokumentnr. 1000000030330 v02	Januar 2019	Informasjon om v2.5-reagenssett lagt til.
Dokumentnr. 1000000030330 v01	August 2018	Oppdaterte regulatoriske merkinger.
Dokumentnr. 1000000030330 v00	November 2017	Første versjon.

Teknisk hjelp

Kontakt teknisk støtte hos Illumina for teknisk hjelp.

Nettsted: www.illumina.com
E-post: techsupport@illumina.com

Telefonnumre til Illuminas kundestøtte

Region	Gratis	Regionalt
Nord-Amerika	+1.800.809.4566	
Australia	1.800.775.688	
Belgia	+32 80077160	+32 34002973
Danmark	+45 80820183	+45 89871156
Finland	+358 800918363	+358 974790110
Frankrike	+33 805102193	+33 170770446
Hongkong	800960230	
Irland	+353 1800936608	+353 016950506
Italia	+39 800985513	+39 236003759
Japan	0800.111.5011	
Kina	400.066.5835	
Nederland	+31 8000222493	+31 207132960
New Zealand	0800 451 650	
Norge	+47 800 16836	+47 21939693
Singapore	+1.800.579.2745	
Spania	+34 911899417	+34 800300143
Storbritannia	+44 8000126019	+44 2073057197
Sveits	+41 565800000	+41 800200442
Sverige	+46 850619671	+46 200883979
Taiwan	00806651752	
Tyskland	+49 8001014940	+49 8938035677
Østerrike	+43 800006249	+43 19286540
Andre land	+44 1799 534 000	

Sikkerhetsdatablad – Tilgjengelige på Illuminas nettsted på support.illumina.com/sds.html.

Produktdokumentasjon – Tilgjengelig for nedlasting i PDF-format fra Illuminas nettsted. Gå til support.illumina.com, velg et produkt, og velg deretter **Documentation & Literature** (Dokumentasjon og litteratur).



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



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

© 2021 Illumina, Inc. Med enerett.

illumina®