

Analysmodulen Somatic Variant (somatisk variant) i Local Run Manager (lokal körningshanterare)

Arbetsflödesguide för NextSeq 550Dx

FÖR IN VITRO-DIAGNOSTISKT BRUK

Översikt	3
Ange körningsinformation	4
Analysmetoder	6
Visa körnings- och provdata	7
Analysrapport	7
Analysutdatafiler	9
Basbestämning och indexmångfald	15
Revisionshistorik	17
Teknisk hjälp	18



Dokumentet och dess innehåll tillhör Illumina, Inc. och dess dotterbolag ("Illumina") och är endast avsett för användning enligt avtal i samband med kundens bruk av produkterna som beskrivs häri. Allt annat bruk är förbjudet. Dokumentet och dess innehåll får ej användas eller distribueras i något annat syfte och/eller återges, delges eller reproduceras på något vis utan föregående skriftligt tillstånd från Illumina. I och med detta dokument överlåter Illumina inte någon licens som hör till dess patent, varumärke eller upphovsrätt, eller i enlighet med rättspraxis eller liknande tredjepartsrättigheter.

Instruktionerna i detta dokument ska följas till punkt och pricka av kvalificerad och lämpligt utbildad personal för att säkerställa rätt och säker produktanvändning i enlighet med beskrivning häri. Hela innehållet i dokumentet ska läsas och förstås i sin helhet innan produkten (produkterna) används.

UNDERLÅTENHET ATT LÄSA OCH FÖLJA ALLA INSTRUKTIONER HÄRI I SIN HELHET KAN MEDFÖRA SKADA PÅ PRODUKTEN/PRODUKTERNA, PERSONSKADA, INKLUSIVE SKADA PÅ ANVÄNDAREN/ANVÄNDARNA ELLER ANDRA PERSONER SAMT SKADA PÅ ANNAN EGENDOM, OCH LEDER TILL ATT EVENTUELL GARANTI FÖR PRODUKTEN/PRODUKTERNA BLIR OGILTIG.

ILLUMINA KAN INTE ÅLÄGGAS NÅGOT ANSVAR SOM UPPKOMMER GENOM FELAKTIG ANVÄNDNING AV PRODUKTERNA SOM BESKRIVS HÄRI (INKLUSIVE DELAR DÄRI ELLER PROGRAM).

© 2021 Illumina, Inc. Med ensamrätt.

Alla varumärken tillhör Illumina, Inc. eller respektive ägare. Specifik varumärkesinformation finns på www.illumina.com/company/legal.html.

Översikt

Modulen Somatic Variant (somatisk variant) i Local Run Manager (lokal körningshanterare) är avsedd att användas tillsammans med analysen Illumina TruSeq Custom Amplicon Kit Dx och NextSeq 550Dx. Vid användning med modulen Somatic Variant (somatisk variant) är analysen avsedd för beredning av bibliotek som används vid sekvensering av DNA från formalinfixerad, paraffinbäddad (FFPE) vävnad. Vid analysen detekteras somatiska mutationer vid låga variantfrekvenser.

Analysmodulen utvärderar korta områden av amplifierat DNA (amplikoner) med avseende på varianter. Fokuserad sekvensering av ampliconer möjliggör hög täckning av särskilda områden i ett stort antal prover. Analysmodulen utför en sekundär analys och genererar rapporter från sekvenseringskörningar med hjälp av en dubbelsträngsmetod som omfattar en framåtriktad och en omvänd oligo-pool. Se bipacksedeln till *TruSeq Custom Amplicon Kit Dx* (dokumentnr 1000000029772).

Analysmodulen Somatic Variant (somatisk variant) kräver förbrukningsmaterial för 300 sekvenseringscykler. Mer information finns i bipacksedeln till *NextSeq 550Dx högproduktivt reagenskit v2* eller *NextSeq 550Dx högproduktivt reagenskit v2.5*.

Om den här handboken

Den här handboken innehåller anvisningar för hur du ställer in körningsparametrar för sekvensering och analys för analysmodulen Somatic Variant (somatisk variant). Information om instrumentpanelen och systeminställningarna i Local Run Manager (lokal körningshanterare) finns i *Referensguide för NextSeq 550Dx-instrument* (dokumentnr 1000000009513).

Visa Local Run Manager (lokal körningshanterare)

Gränssnittet för Local Run Manager (lokal körningshanterare) visas i NextSeq 550Dx systemprogramvara (NOS) eller via en webbläsare. Chromium är den webbläsare som stöds.



OBS!

Om du använder en webbläsare som inte stöds hämtar du den kompatibla webbläsaren när du uppmanas till det i meddelandet "Confirm Unsupported Browser" (bekräfta webbläsare som inte stöds). Välj **här** för att hämta den version av Chromium som stöds.

Visa på instrumentets skärm

- 1 Välj ett av följande alternativ om du vill visa gränssnittet för Local Run Manager på instrumentets skärm:
 - ▶ Gå till startskärmen i NOS och välj **Local Run Manager** (lokal körningshanterare). Klicka på X i övre högra hörnet för att återgå till NOS när du är klar.
 - ▶ Välj ikonen Minimize NOS (Minimera NOS), öppna Chromium-webbläsaren som finns på instrumentet och ange **http://localhost** i adressfältet. Endast administratörsanvändare kan minimera NOS.

Visa från en nätverksansluten dator

- 1 Öppna en Chromium-webbläsare på en dator med tillgång till samma nätverk som instrumentet och anslut med hjälp av instrumentets IP-adress eller instrumentnamn. Till exempel **http://myinstrument**.

Ange körningsinformation

Ställa in parametrar

- 1 Logga in i Local Run Manager (lokal körningshanterare).
- 2 Välj **Create Run** (skapa körning) och sedan **Somatic Variant** (somatisk variant).
- 3 Ange ett namn på körningen som identifierar körningen från sekvensering till analys. Använd alfanumeriska tecken, blanksteg, understreck eller bindestreck.
- 4 **[Valfritt]** Ange en beskrivning av körningen som hjälper till att identifiera körningen. Använd alfanumeriska tecken, blanksteg, understreck eller bindestreck.
- 5 Välj antal prover och indexset i listrutan.
Överväg följande information när du väljer ett alternativ.
 - ▶ Listrutan visar antal prover och ett indexset. Exempelvis indikerar 24-Set 1 att 24 prover ska testas, med index från indexset 1.
 - ▶ Indexsetnumren hänvisar till olika set av i5-index. Set 1 och Set 2 ger båda indexmångfald. Två indexset erbjuds för att förhindra att ett enda set förbrukas.
 - ▶ Välj antalet prover som är närmast det antal prov du testas. Om det exakta antalet prov inte finns med i listan väljer du det närmaste antalet som innehåller färre prover än det antal som testas. Om du exempelvis vill testa 18 prover väljer du 16 prover.
 - ▶ Provbrunnar och indexkombinationer som uppfyller kraven för indexmångfald är markerade med grönt. Om andra brunnar och indexkombinationer väljs visas ett meddelande om att kraven för indexmångfald inte uppfylls när körningen sparas.

Importerera manifestfiler för körningen

- 1 Kontrollera att de manifest du vill importera finns på en åtkomlig plats i nätverket eller på en USB-enhet.
- 2 Välj **Import Manifests** (importera manifest).
- 3 Navigera till manifestfilen och välj de manifest som du vill lägga till.



OBS!

Lägg till manifesten via funktionen Module Settings (modulinställningar) för att göra manifestfilerna tillgängliga för alla körningar med analysmodulen Somatic Variant (somatisk variant). Den här funktionen kräver behörighet på administratörsnivå. Mer information finns i *Referensguide för NextSeq 550Dx-instrument* (dokumentnr 1000000009513).


Ange prover för körningen

Ange prover för körningen med hjälp av ett av alternativen och anvisningarna som följer.


- ▶ **Enter samples manually** (ange prover manuellt) – Använd den tomma tabellen på skärmen Create Run (skapa körning).
- ▶ **Import samples** (importera prover) – navigera till en extern fil i ett format med kommateckenavgränsade fält (*.csv). En mall kan hämtas på skärmen Create Run (skapa körning).

När du har fyllt i provtabellen kan du exportera provinformationen till en extern fil. Använd filen som referens när du bereder bibliotek eller importerar filen för en annan körning.

Ange prover manuellt

- 1 Ange ett unikt provnamn i fältet Sample Name (provnamn).
Använd alfanumeriska tecken, bindestreck eller understreck.
Provnamnet fylls automatiskt i för motsvarande brunn i den andra poolen.
- 2 **[Valfritt]** Högerklicka och välj kontrolltyp för positiva eller negativa kontrollprover.
Kontrollen i en provbrunn fylls automatisk i motsvarande brunn i den andra poolen med samma kontroll.
- 3 **[Valfritt]** Ange en beskrivning av provet i fältet Sample Description (beskrivning av provet).
Använd alfanumeriska tecken, bindestreck eller understreck.
Provets beskrivning fylls automatiskt i för motsvarande brunn i den andra poolen.
Beskrivningar av prov är kopplade till ett prov-ID. Beskrivningar av prov skrivs över om samma prov-ID används igen i en senare körning.
- 4 Välj en Index 1-adapter i listrutan Index 1 (i7).
Om du använder de föreslagna provbrunnarna fyller programvaran automatiskt i i7- och i5-indexadapterar som uppfyller mångfaldskraven för index. Om det exakta antalet prov du testat inte finns med i listan måste du komma ihåg att välja indexadapterar för extra brunnar. Om du behöver välja index för extra brunnar eller inte använder de rekommenderade kombinationerna av indexadapterar bör du se till att läsa [Basbestämning och indexmångfald på sidan 15](#) innan du väljer index.
- 5 Välj en Index 2-adapter i listrutan Index 2 (i5).
- 6 Välj en manifestfil i listrutan Manifest.
Proverna i pool A måste ha ett annat manifest än proverna i pool B.
- 7 Välj ett alternativ för att visa, skriva ut eller spara plattlayouten som referens för beredning av bibliotek:
 - ▶ Välj ikonen  **Print** (skriv ut) för att visa plattlayouten. Välj **Print** (skriv ut) för att skriva ut plattlayouten.
 - ▶ Välj **Export** (exportera) för att exportera provinformation till en extern fil.Se till att manifest- och provinformationen stämmer. Felaktig information kan påverka resultatet.
- 8 Välj **Save Run** (spara körning).

Importera prover

- 1 Välj **Import Samples** (importera prover) och bläddra till platsen där filen med provinformation finns.
Två typer av filer kan importeras.
 - ▶ Välj **Template** (mall) på skärmen Create Run (skapa körning) för att skapa en ny plattlayout. Mallfilen innehåller rätt kolumnrubriker för import. Ange provinformation i varje kolumn för proverna i körningen. Ta bort exempelinformation i oanvända celler och spara sedan filen.
 - ▶ Använd en fil med provinformation som exporterats från modulen Somatic Variant (somatisk variant) med exportfunktionen.
- 2 Välj ikonen  **Print** (skriv ut) för att visa plattlayouten.
- 3 Välj **Print** (skriv ut) för att skriva ut plattlayouten som referens vid beredning av bibliotek.
- 4 **[Valfritt]** Välj **Export** (exportera) för att exportera provinformation till en extern fil.
Se till att manifest- och provinformationen stämmer. Felaktig information kan påverka resultatet.
- 5 Välj **Save Run** (spara körning).

Redigera en körning

Information om hur informationen i körningen kan redigeras före sekvensering finns i *Referensguide för NextSeq 550Dx-instrument (dokumentnr 100000009513)*.

Analysmetoder

Analysmodulen Somatic Variant (somatisk variant) utför följande analyssteg och skriver sedan analysutdatafiler till mappen Alignment (inpassning).

- ▶ Demultiplexerar indexläsningar
- ▶ Genererar FASTQ-filer
- ▶ Passar in mot en referens
- ▶ Identifierar varianter

Demultiplexning

Vid demultiplexning jämförs varje indexläsningssekvens med de indexsekvenser som angetts för körningen. Inga kvalitetsvärden beaktas i det här steget.

Indexläsningar identifieras med följande steg:

- ▶ Prover numreras i den ordning de anges för körningen, med början på 1.
- ▶ Provnummer 0 är reserverat för kluster som inte tilldelats till något prov.
- ▶ Klustren tilldelas till ett prov när indexsekvensen matchar exakt eller när det finns högst en enskild felparning per indexläsning.

Generering av FASTQ-filer

Efter demultiplexningen genererar programvaran intermediära analysfiler i FASTQ-format, vilket är ett textformat som används för att representera sekvenser. FASTQ-filer innehåller läsningar för varje prov och tillhörande kvalitetspoäng. Kluster som inte passerat filtret utesluts.

Varje FASTQ-fil innehåller endast läsningar för ett prov och namnet på det provet ingår i FASTQ-filens namn. FASTQ-filer är inpassningens primära indata. Åtta FASTQ-filer genereras per prov per oligo-pool, fyra från läsning 1 och fyra från läsning 2, vilket ger totalt 16 FASTQ-filer per prov.

Inpassning

Under inpassningssteget passar Smith-Waterman-bandalgoritmen in kluster från varje prov mot amplikonsekvenser som specificerats i manifestfilen.

Smith-Waterman-bandalgoritmen utför semiglobal sekvensinpassning för att fastställa liknande regioner mellan två sekvenser. Smith-Waterman-algoritmen jämför inte hela sekvensen utan jämför i stället segment av alla möjliga längder.

Varje paired end-läsning utvärderas med avseende på dess inpassning till de relevanta sondsekvenserna för läsningen.

- ▶ Läsning 1 utvärderas mot det omvända komplementet hos nedströms lokusspecifika oligonukleotider (Downstream Locus-Specific Oligos – DLSO).
- ▶ Läsning 2 utvärderas mot uppströms lokusspecifika oligonukleotider (ULSO).

- ▶ Om början på en läsning matchar en sondsekvens med högst tre skillnader (felparningar eller förskjutningar på grund av indels) passas hela avläsningens längd in mot amplitonmålet för den sekvensen.
- ▶ Indels inom DLSO och ULSO observeras inte på grund av analysmetoden.

Inpassningar filtreras från inpassningsresultat baserat på felparningsfrekvensen över antingen intresseområdet eller den fullständiga amplitonen, beroende på amplitonens längd. Filtrerade inpassningar anges som ej inpassade i inpassningsfiler och används inte vid variantbestämning.

Variantbestämning

Variantbestämningsprogrammet Pisces som utvecklats av Illumina identifierar varianter som förekommer i låg frekvens i DNA-provet.

Variantbestämningsprogrammet Pisces identifierar SNV:er, MNV:er och små indels i tre steg:

- ▶ överväger varje position i referensgenomet separat
- ▶ räknar baser vid den givna positionen för inpassade läsningar som överlappar positionen
- ▶ beräknar en variantpoäng som mäter bestämningens kvalitet med hjälp av Poisson-modellen. Varianter med en kvalitetspoäng under Q30 utesluts.

Varianter bestäms först för varje pool separat. Därefter jämförs varianterna från vardera pool och kombineras till en enda utdatafil. Om en variant förekommer i båda poolerna och passerar alla filter som anges i *VCF-filanteckningar* på sidan 12 märks varianten som PASS (godkänd) i variantbestämningsfilen (VCF).

Visa körnings- och provdata

- 1 Klicka på körningens namn från instrumentpanelen i Local Run Manager (lokal körningshanterare).
- 2 Gå till fliken Run Overview (körningsöversikt) och granska värdena från sekvenseringskörningen.
- 3 **[Valfritt]** Klicka på ikonen **Copy to Clipboard**  (kopiera till urklipp) för att kopiera sökvägen till körningens utdatamapp.
- 4 Klicka på fliken Sequencing Information (sekvenseringsinformation) för att granska körningsparametrar och information om förbrukningsmaterial.
- 5 Klicka på fliken Samples and Results (prover och resultat) för att visa platsen för analysrapporten.
 - ▶ Om analysen upprepats expanderar du listrutan Select Analysis (välj analys) och väljer rätt analys.
- 6 Klicka på ikonen **Copy to Clipboard**  (kopiera till urklipp) för att kopiera sökvägen till mappen Analysis (analys).

Mer information om flikarna Run Overview (körningsöversikt) och Sequencing Information (sekvenseringsinformation) finns i *Referensguide för NextSeq 550Dx-instrument (dokumentnr 1000000009513)*.

Analysrapport

Analysresultaten sammanfattas på fliken Samples and Results (prov och resultat) och visas i form av en samlingsrapport i mappen Alignment (inpassning). En rapport för varje prov finns även i PDF-filformat för varje prov.

Information på fliken Samples and Results (prov och resultat)

- 1 Klicka på ett prov i listan för att se rapporten för provet.

Tabell 1 Information om körningar och prover

Kolumnrubrik	Beskrivning
Run Status (körningsstatus)	Anger om sekvenseringskörningen lyckades eller misslyckades.
Total Yield (GB) (totalt utbyte (GB))	Antal baser som bestämts i sekvenseringskörningen. Visar gränsen för godkänt samt status (godkänt eller misslyckat).
% ≥ Q30	Procentandelen läsningar i sekvenseringskörningen med en kvalitetspoäng på 30 (Q30) eller högre. Visar gränsen för godkänt samt status (godkänt eller misslyckat).
Sample Name (provnamn)	Det provnamn som angivits när körningen skapades.
Total PF Reads (totalt antal godkända läsningar)	Totalt antal läsningar som passerat filtren.
Read 1% ≥ Q30 (läsning 1 % ≥ Q30)	Procentandel läsningar i läsning 1 med en kvalitetspoäng på 30 (Q30) eller högre för provet.
Read 2% ≥ Q30 (läsning 2 % ≥ Q30)	Procentandel läsningar i läsning 2 med en kvalitetspoäng på 30 (Q30) eller högre för provet.
Autosome Call Rate (bestämningsfrekvens för autosomer)	Antalet genomiska positioner i autosomerna (kromosom 1 till och med 22) som uppfyller ett fördefinierat konfidenströskelvärde, delat med det totala antalet autosomala genomiska positioner som undersökts. Bestämningsfrekvensen beskrivs per prov och rapporteras som en procentandel som beräknas som 1 minus antalet autosomala positioner med ofullständiga bestämningar delat med det totala antalet autosomala positioner som sekvenserats.

Tabell 2 Information i provrapporten

Kolumnrubrik	Beskrivning
Sample (prov)	Det provnamn som angivits när körningen skapades.
Report Date (rapportdatum)	Datum när rapporten genererats.
Sample Information (provinformation)	Det prov-ID som angavs när körningen skapades, totalt antal läsningar som passerat filtret i provet, procentandel läsningar för provet med en kvalitetspoäng på 30 (Q30) eller högre samt autosomal bestämningsfrekvens.
Amplicon Summary (amplikonsammanfattning)	Det totala antalet ampliconområden som sekvenserats och den totala längden, i baspar, för sekvenserade ampliconer i målområdena, för provet i pool A och pool B och för manifestfilen som används i vardera pool. I manifestfilen specificeras det referensgenom och de målreferensområden som använts i inpassningssteget.
Read Level Statistics (statistik på läsningsnivå)	Antal och procentandel av läsningarna för provet som täcker varje position i referensen, för läsning 1 och läsning 2 i pool A och pool B.
Variants Summary (variantsammanfattning)	Antal SNV:er, insertioner och deletioner som detekterats för provet och som passerat de föreslagna värdena för att avgöra om kvalitetsresultaten ligger inom ett acceptabelt intervall.
Coverage Summary (täckningssammanfattning)	Det totala antalet inpassade baser delat med målområdets storlek, samt procentandelen ampliconområden med täckningsvärden över den lägre täckningströskeln på 0,2 * genomsnittlig amplicontäckning, för provet i pool A och pool B.

Kolumnrubrik	Beskrivning
Coverage Plots (täckningsdiagram)	Diagrammet Coverage by Amplicon Region (täckning per ampliconområde) visar täckningen i olika ampliconområden för provet. Områden med lägre täckningsvärden än täckningströskelvärdet är markerade med rött. Genomsnittet för alla värden anges med en orange linje. Ett diagram visar täckningen för pool A och pool B.
Software Versions (programvaruversioner)	Programvaruversioner då provet sekvenserades. Omfattar versionerna för NextSeq 550Dx systemprogramvara (NOS), programvaran Local Run Manager (lokal körningshanterare), programvaran RTA och modulen Somatic Variant (somatisk variant).

Analysutdatafiler

Följande analysutdatafiler genereras för analysmodulen Somatic Variant (somatisk variant) och ger analysresultat för inpassning och variantbestämning. Analysutdatafiler finns i mappen Alignment (inpassning).

Filnamn	Beskrivning
Demultiplexning (*.txt)	Intermediära filer som innehåller sammanfattande demultiplexningsresultat.
FASTQ (*.fastq.gz)	Intermediära filer som innehåller basbestämningar med kvalitetspoäng. FASTQ-filer är inpassningsstegets primära indata.
Inpassningsfiler i BAM-format (*.bam)	Innehåller inpassade läsningar för ett givet prov.
Variantbestämningsfiler per pool i VCF-format (*.vcf.gz)	Innehåller bestämda varianter i varje position från antingen den framåtriktade poolen eller den omvända poolen.
Variantbestämningsfiler i Genome VCF-format (*.genome.vcf.gz)	Innehåller genotypen för varje position, oavsett om den bestämts som variant eller som referens.
Sammanställda variantbestämningsfiler i VCF-format (*.vcf.gz)	Innehåller bestämda varianter i varje position från båda poolerna.
AmpliconCoverage_M1.tsv	Innehåller information om täckning per amplicon per prov för varje manifest som anges. M-numret står för manifestets nummer.

Filformat för demultiplexning

Under demultiplexningsprocessen läses den indexsekvens som är kopplad till varje kluster för att avgöra från vilket prov klustret kommer. Mappningen mellan kluster och provnummer skrivs till en demultiplexningsfil (*.demux) för varje platta på flödescellen.

Demultiplexningsfilens namngivningsformat är **s_1_X.demux**, där X är plattans nummer.

Demultiplexningsfiler börjar med en rubrik:

- ▶ version (4-byte-heltal), för närvarande 1
- ▶ klusterantal (4-byte-heltal).

Resten av filen består av provnumren för varje kluster från plattan.

När demultiplexningsteget avslutats genererar programvaran en demultiplexningsfil med namnet **DemultiplexSummaryF1L1.txt**.

- ▶ I filnamnet står **F1** för flödescellens nummer.
- ▶ I filnamnet står **L1** för spårets nummer.
- ▶ Demultiplexningen resulterar i en tabell med 1 rad per platta och 1 kolumn per prov, medräknat prov 0.

- ▶ De vanligast förekommande sekvenserna i indexläsningar.

FASTQ-filformat

FASTQ är ett textbaserat filformat som innehåller basbestämning och kvalitetsvärden per läsning. Varje post innehåller fyra rader:

- ▶ identifieraren
- ▶ sekvensen
- ▶ ett plustecken (+)
- ▶ Phred-kvalitetspoäng i ett ASCII + 33-kodat format.

Identifieraren är formaterad som:

@Instrument:Körnings-ID:Flödescell-ID:Spår:Platta:X:Y Läsningsnummer:Filterflagga:0:Provnummer.

Exempel:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAAA9#:<#<;<<<????#=#
```

BAM-filformat

En BAM-fil (*.bam) är den komprimerade binära versionen av en SAM-fil som används för att representera inpassade sekvenser på upp till 128 Mb. SAM- och BAM-format beskrivs i detalj i samtools.github.io/hts-specs/SAMv1.pdf.

För BAM-filer används namngivningsformatet **Provnamn_S#.bam**. # är provnumret som bestäms av den ordning proverna angetts i körningen.

BAM-filer innehåller ett rubrikavsnitt och ett inpassningsavsnitt:

- ▶ **Header** (rubrik) – innehåller information om hela filen, exempelvis provnamn, provlängd och inpassningsmetod. Inpassningar i inpassningsavsnittet är kopplade till specifik information i rubrikavsnittet.
- ▶ **Alignments** (inpassningar) – innehåller läsningens namn, läsningens kvalitet, inpassningsinformation och anpassade taggar. Läsningens namn inbegriper kromosom, startkoordinat, inpassningskvalitet och strängen för parningsbeskrivning.

Inpassningsavsnittet innehåller följande information för varje läsning eller läspar:

- ▶ **AS:** paired end-inpassningskvalitet.
- ▶ **BC:** streckkodstagg som visar det demultiplexade prov-ID som är kopplat till läsningen.
- ▶ **SM:** single end-inpassningskvalitet.
- ▶ **XC:** sträng med parningsbeskrivning.
- ▶ **XN:** tagg för amplikonnamn som registrerar det amplikon-ID som är kopplat till läsningen.

BAM-indexfiler (*.bam.bai) innehåller en förteckning över motsvarande BAM-fil.

VCF-filformat

Variant Call Format (VCF) är ett vanligt filformat som utvecklats av forskarsamfundet inom genetik. Det innehåller information om varianter som finns på särskilda positioner i ett referensgenom. VCF-filer har ändelsen .vcf

VCF-filens rubrik innehåller VCF-filformatets version och variantbestämningsprogrammets version samt anger de anteckningar som används i resten av filen. I rubriken ingår även referensgenomfilen och BAM-filen. Den sista raden i rubriken innehåller kolumnrubrikerna för dataraderna. Varje datarad i VCF-filen innehåller information om en variant.

VCF-filrubriker

Rubrik	Beskrivning
CHROM	Kromosom i referensgenomet. Kromosomer visas i samma ordning som i FASTQ-referensfilen.
POS	Variantens enbasposition på referenskromosomen. För SNP:er är den här positionen referensbasen hos varianten. För indels eller deletioner är den här positionen referensbasen direkt före varianten.
ID	Variantens rs-nummer som erhållits från dbSNP.txt, om tillämpligt. Om det finns flera rs-nummer på denna position är listan semikolonavgränsad. Om ingen dbSNP-post finns på den här positionen används en markör för saknat värde ('.') .
REF	Referensgenotypen. En deletion av ett enda T representeras exempelvis som referens TT och alternativ T. En nukleotidvariant där A ersatts av T representeras som referens A och alternativ T.
ALT	De alleler som skiljer sig från referensläsningen. Exempelvis representeras en insertion av ett enda T som referens A och alternativ AT. En nukleotidvariant där A ersatts av T representeras som referens A och alternativ T.
QUAL	En Phred-kvalitetspoäng som tilldelats av variantbestämningsprogrammet. Högre poäng anger högre konfidensnivå för varianten och lägre sannolikhet för fel. För kvalitetspoängen Q är den beräknade sannolikheten för fel $10^{-(Q/10)}$. Exempelvis har uppsättningen av Q30-bestämningar en felfrekvens på 0,1 %. Många variantbestämningsprogram tilldelar kvalitetspoäng utifrån sina statistiska modeller, som är höga i förhållande till den observerade felfrekvensen.

VCF-filanteckningar

Rubrik	Beskrivning
FILTER	<p>Om alla filter passerar skrivs PASS (godkänd) i filterkolumnen.</p> <ul style="list-style-type: none"> • LowDP – gäller platser med täckningsdjup under 450x i någon av poolerna. För amplikonpositioner som täcks av både framåtriktad och en omvänd läsning motsvarar det här 900x täckning vid enkel läsning. • LowGQ – genotypbestämningens kvalitet (GQ) ligger under cutoff. • q30 – kvalitetspoäng < 30. • LowVariantFreq – variantfrekvensen är lägre än det angivna tröskelvärdet. • PB – bias för sondpool. Varianten hittades inte eller hittades med låg frekvens i en eller två sondpools. • R3x6 – antalet upprepningar (1 till 3 bp långa) intill variantbestämningarna ≥ 6. • SB – strängbias är större än det givna tröskelvärdet.
INFO	<p>Följande är möjliga poster i INFO-kolumnen:</p> <ul style="list-style-type: none"> • AC – allelantal i genotyper för varje ALT-allel, i samma ordning som de visas. • AF – allelfrekvens för varje ALT-allel, i samma ordning som de visas. • EN – det totala antalet alleler i bestämda genotyper. • CD – en flagga som anger att SNP förekommer i kodningsområdet på minst en RefGene-post. • DP – djup (antal basbestämningar som passats in på en position och används i variantbestämning). • Exon – en kommaseparerad lista över exonområden som lästs från RefGene. • FC – funktionell konsekvens. • GI – en kommaseparerad lista över gen-ID:n som lästs från RefGene. • QD – variantkonfidens/kvalitet efter djup. • TI – en kommaseparerad lista över transkript-ID:n som lästs från RefGene.
FORMAT	<p>I formatkolumnen visas fält separerade med kolon, exempelvis GT:GQ. Listan över fält beror på vilket variantbestämningssystem som använts. Följande är tillgängliga fält:</p> <ul style="list-style-type: none"> • AD – inmatning av X och Y, där X är antalet referensbestämningar och Y är antalet alternativa bestämningar. • DP – ungefärligt läsningsdjup. Läsningar med MQ = 255 eller med felparningar filtreras bort. • GQ – genotypkvalitet. • GQX – genotypkvalitet. GQX är det mindre av GQ-värdet och QUAL-kolumnen. De här värdena är i allmänhet likartade, men genom att använda det minsta värdet blir GQX det mest konservativa måttet på genotypkvalitet. • GT – genotyp. 0 motsvarar referensbasen, 1 motsvarar den första posten i ALT-kolumnen och så vidare. Snedstreck (/) anger att ingen fasningsinformation finns tillgänglig. • NL – brusnivå, en uppskattning av basbestämningens brus vid den här positionen. • PB – bias för sondpool. Värdet närmare 0 indikerar mer bias mot en sondpool och lägre konfidensnivå för en variantbestämning. • SB – strängbias på den här positionen. Större negativa värden indikerar mindre bias medan värden nära 0 visar på större bias. • VF – variantfrekvens, procentandel av läsningarna som stöder den alternativa allelen.
SAMPLE (prov)	I provkolumnen anges de värden som specificerats i FORMAT-kolumnen.

Genome VCF-filer

Filer av typen Genome VCF (gVCF) är VCF v 4.1-filer som följer en uppsättning konventioner för att representera alla platser i genomet i ett relativt kompakt format. gVCF-filerna (*.genome.vcf.gz) innehåller alla platser i intresseområdet i en enda fil för varje prov.

gVCF-filen visar saknade bestämningar på positioner som inte passerat alla filter. En genotyptagg (GT) på ./ anger en saknad bestämning.

Mer information finns på sites.google.com/site/gvcftools/home/about-gvcf.

VCF-filer, per pool och sammanställda

Arbetsflödet i Somatic Variant (somatisk variant) genererar 2 uppsättningar av variantbestämningsfiler:

- ▶ **VCF-filer per pool** – innehåller varianter som bestämts i antingen den framåtriktade eller den omvända poolen. Filerna per pool skrivs till mappen VariantCallingLogs.
- ▶ **Sammanställda VCF-filer** – innehåller varianter som bestämts från båda poolerna. De sammanställda filerna skrivs till mappen Alignment (inpassning).

VCF-filer per pool och sammanställda VCF-filer omfattar både VCF-filer (*.vcf) och gVCF-filer (*.genome.vcf) och följande namngivningsregler används, där S# representerar den ordning som provet visas för körningen.

- ▶ **Rapporter för alla platser** – SampleName_S#.genome.vcf
- ▶ **Rapporter för endast varianter** – SampleName_S#.vcf

Programvaran jämför VCF-filerna per pool och kombinerar data vid varje position för att skapa en sammanställd VCF-fil för provet.

Variantbestämningar från varje pool sammanfogas till sammanställda VCF-filer med hjälp av följande kriterier.

Kriterier	Resultat
En referensbestämning i varje pool.	Referensbestämning
En referensbestämning i en pool och en variantbestämning i den andra poolen.	Filtrerad variantbestämning
Matchande variantbestämningar med liknande frekvenser i varje pool.	Variantbestämning
Matchande variantbestämningar med påtagligt olika frekvenser i varje pool.	Filtrerad variantbestämning
Variantbestämningar som inte matchar i varje pool.	Filtrerad variantbestämning

Mått från båda poolerna sammanfogas med hjälp av värdena nedan.

Mått	Värde
Djup	Summan av djupen från båda poolerna.
Variantfrekvens	Totalt antal varianter delat med totalt täckningsdjup.
Q-Score (kvalitetspoäng)	Minimivärde för båda poolerna.

Amplikontäckningsfil

En amplikontäckningsfil genereras för varje manifestfil. M-numret i filnamnet representerar manifestets nummer.

Varje fil innehåller en rubrikrad som inbegriper alla prov-ID:n som är kopplade till manifestet. Filen innehåller följande information:

- ▶ mål-ID som det anges i manifestet.
- ▶ täckningsdjupet för läsningar som passerar filtret.

Kompletterande utdatafiler

Följande utdatafiler ger kompletterande information eller sammanfattar körningsresultat och analysfel. Även om filerna inte krävs för bedömning av analysresultaten kan de användas i felsökningssyfte. Alla filer finns i mappen Alignment (inpassning) om inget annat anges.

Filnamn	Beskrivning
AnalysisLog.txt	Bearbetningslogg som beskriver varje steg som vidtagits under analysen av den aktuella körningsmappen. Filen innehåller inga felmeddelanden. Finns i mappen Alignment (inpassning).
AnalysisError.txt	Bearbetningslogg som visar en lista över alla fel som inträffat under analysen. Filen är tom om inga fel inträffat. Finns i mappen Alignment (inpassning).
DemultiplexSummaryF1L1#.txt	Rapporterar demultiplexningens resultat i en tabell med 1 rad per platta och 1 kolumn per prov. Numren representerar spår 1, 2, 3 eller 4 i flödescellen. Finns i mappen Alignment (inpassning).
AmpliconRunStatistics.xml	Innehåller kortfattad statistik som är specifik för körningen. Finns i mappen Alignment (inpassning).

Analysmapp


Analysmappen innehåller filer som genererats av programvaran Local Run Manager (lokal körningshanterare). Förhållandet mellan utdatamappen och analysmappen kan sammanfattas som följer:

- ▶ Under sekvenseringen fyller programvaran för realtidsanalys (RTA) utdatamappen med filer som genererats vid bildanalys, basbestämning och kvalitetsbedömning.
- ▶ RTA kopierar filer till analysmappen i realtid. När RTA gett en kvalitetspoäng till varje bas i varje cykel skriver programvaran filen RTAComplete.txt till båda mapparna.
- ▶ Analysen börjar så fort filen RTAComplete.txt finns i mapparna.
- ▶ Under analysens gång skriver Local Run Manager (lokal körningshanterare) utdatafiler till analysmappen och kopierar sedan filerna tillbaka till utdatamappen.

Inpassningsmappar

Varje gång som analysen ställs tillbaka i kö skapar Local Run Manager (lokal körningshanterare) en inpassningsmapp med namnet **Alignment_N** (inpassning_N), där N är ett löpnummer.

Mappstruktur

 **Alignment** (inpassning) – innehåller filer av typerna *.bam, *.vcf och FASTQ samt filer som är specifika för analysmodulen.

 **Date and Time Stamp (datum- och tidsstämpel)** – datum- och tidsstämpel för analysen i formatet ÅÅÅÅMMDD_HHMMSS

-  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
- 📄 Sample1.summary.csv
- 📄 Sample1.vcf.gz
- 📄 Sample1.bam
- 📁 **FASTQ**
 - 📁 **Sample1** (prov1)
 - 📄 Sample1_L001_R1_001_fastq.gz
 - 📁 **Stats** (statistik)
 - 📄 DemuxSummaryF1L1.txt
 - 📄 FastqSummaryF1L1.txt
- 📁 **Data**
 - 📁 **Intensities** (intensiteter)
 - 📁 **BaseCalls** (basbestämningar)
 - 📁 L001 – innehåller *.bcl-filer.
 - 📁 L001 – innehåller *.locs-filer.
 - 📁 **RTA Logs** (RTA-loggar) – innehåller loggfiler från RTA-programvaruanalys.
 - 📁 **InterOp** – innehåller binärfiler som används för att rapportera värden från sekvenseringskörning.
 - 📁 **Logs** (loggar) – innehåller loggfiler som beskriver åtgärder som utförs under sekvensering.
 - 📄 RTAComplete.txt
 - 📄 RunInfo.xml
 - 📄 RunParameters.xml

Basbestämning och indexmångfald

När prover sekvenseras i NextSeq 550Dx-instrumentet bestäms via basbestämning en bas (A, C, G eller T) för varje kluster på en given platta, eller ett bildområde på flödescellen, i en viss cykel. I NextSeq 550Dx-instrumentet används tvåkanalssekvensering, som endast kräver två bilder för att koda data för fyra DNA-baser, en från den röda kanalen och en från den gröna kanalen.

Processen för basbestämning vid indexläsning skiljer sig från basbestämning vid andra läsningar.

Indexläsningar måste börja med minst en annan bas än G i någon av de första två cyklerna. Om en indexläsning börjar med två basbestämningar av G genereras inte någon signalintensitet. Signalen måste finnas i någon av de första två cyklerna för att demultiplexningen ska fungera.

Vid val av index när körningen skapas visas en varning om låg mångfald ifall index inte uppfyller kraven för mångfald. Undvik varningen om låg mångfald genom att välja indexsekvenser som ger signal i båda kanalerna för varje cykel.

- ▶ Röd kanal: A eller C.
- ▶ Grön kanal: A eller T.

Den här processen för basbestämning säkerställer att resultatet blir rättvisande vid analys av lågplex-prover. Mer information om sekvenserna för dina index finns i bipacksedeln till *TruSeq Custom Amplicon Kit Dx* (dokumentnr 1000000029772).

När du skapar körningen i Local Run Manager (lokal körningshanterare) väljer du antalet prover som ska testas. Föreslagna indexkombinationer som uppfyller kraven för indexmångfald fylls automatiskt i av programmet. De är inte nödvändigt att använda de föreslagna indexkombinationerna men det rekommenderas.

Revisionshistorik

Dokument	Datum	Ändringsbeskrivning
Dokumentnr 1000000030330 v04	Augusti 2021	Adressen till den auktoriserade europeiska representanten har uppdaterats.
Dokumentnr 1000000030330 v03	April 2020	Adressen till den auktoriserade europeiska representanten har uppdaterats. Adressen till den australiska sponsorn har uppdaterats.
Dokumentnr 1000000030330 v02	Januari 2019	Lade till information om v2.5-reagenskit.
Dokumentnr 1000000030330 v01	Augusti 2018	Uppdaterad reglerande märkning.
Dokumentnr 1000000030330 v00	November 2017	Första version.

Teknisk hjälp

Kontakta Illuminas tekniska support för all form av teknisk hjälp.

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

Telefonnummer till Illuminas kundtjänst

Region	Avgiftsfritt	Lokalt
Nordamerika	+1 800-8094566	
Australien	+1 800-775688	
Belgien	+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
Italien	+39 800985513	+39 236003759
Japan	0800-1115011	
Kina	400-0665835	
Nederländerna	+31 8000222493	+31 207132960
Norge	+47 800 16836	+47 21939693
Nya Zeeland	0800-451650	
Österrike	+43 800006249	+43 19286540
Schweiz	+41 565800000	+41 800200442
Singapore	+1 800-5792745	
Spanien	+34 911899417	+34 800300143
Storbritannien	+44 8000126019	+44 2073057197
Sverige	+46 850619671	+46 200883979
Taiwan	00806651752	
Tyskland	+49 8001014940	+49 8938035677
Övriga länder	+44 1799-534000	

Säkerhetsdatablad (SDS) – Finns på Illuminas webbplats på support.illumina.com/sds.html.

Produktdokumentation – Kan hämtas i PDF-format på Illuminas webbplats. Gå till support.illumina.com, välj en produkt och klicka sedan på **Dokumentation och litteratur**.



Illumina
5200 Illumina Way
San Diego, California 92122 USA
+1 800-8094566
+1 858-2024566 (utanför Nordamerika)
techsupport@illumina.com
www.illumina.com

CE



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

Australisk sponsor

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

FÖR IN VITRO-DIAGNOSTISKT BRUK

© 2021 Illumina, Inc. Med ensamrätt.

illumina®