

illumina®

TruSight Whole Genome Analysis Application

Produktdokumentation

TILLHÖR ILLUMINA

Dokumentnr 200049931 v00

April 2024

FÖR IN VITRO-DIAGNOSTISKT BRUK.

Revisionshistorik

Dokument	Datum	Beskrivning av ändring
Dokumentnr 200049931 v00	April 2024	Första utgåvan.

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 noggrant och uttryckligen av kvalificerad och lämpligt utbildad personal för att säkerställa rätt och säker produktanvändning i enlighet med beskrivningen 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).

© 2024 Illumina, Inc. Med ensamrätt.

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

Innehållsförteckning

Revisionshistorik	ii
Översikt	1
Komma igång	1
Inställningar	2
Skapande av körning	4
Körningsrevision	5
Repetera analys	6
Visa körning och resultat	7
Sammanfattning av utdatafil	8
Information om QC-rapport	8
Information om variantbestämningsfil	12
FASTQ-filer	20
CRAM-filer	21
Teknisk hjälp	22

Översikt

TruSight Whole Genome Analysis Application används för att planera sekvenseringskörningar för TruSight Whole Genome och automatiskt initiera analys efter att körningen har slutförts. Analysen inkluderar demultiplexering, FASTQ-generering, avläsningsmappning, inpassning mot det grafaktiverade humana referensgenomet GrCh38/hg38 och variantbestämning med användning av Illumina DRAGEN-server för NovaSeq 6000Dx.

I olika stadier av analysarbetsflödet utför applikationen kvalitetskontroll (QC) enligt definierade sekvenserings-, FASTQ- och provbiblioteksmått och genererar rapporter med resultaten. För prover som klarar alla kvalitetskontrollsteg genererar programmet stödfile för användning i nedströms könscellsapplikationer.

TruSight Whole Genome Analysis Application kör DRAGEN variantbestämmare, inklusive Small Variant Caller (bestämning av liten variant), Copy Number Variant Caller (bestämning av kopietalsvariation) (CNV) och Repeat Expansion Detection (upprepad expansionsdetektering) med ExpansionHunter.

Applikationen utför också annoteringar av låg, mellanliggande eller hög konfidensnivå för små varianter och inkluderar denna annotering i utdatafilen.

Komma igång

Se till att TruSight Whole Genome Analysis Application är installerad på NovaSeq 6000Dx-instrumentet som ska användas för sekvensering som en del av TruSight Whole Genome. Installerade program finns på skärmen Applications (Program) på NovaSeq 6000Dx Instrument eller i Illumina Run Manager med en webbläsare på en nätverksansluten dator. Kontakta din lokala Illumina-fältrepresentant för hjälp med att schemalägga installationen.

Lagringskrav för data

Se NovaSeq 6000Dx produktdokumentation (dokument nr 200010105) Illumina DRAGEN Server för NovaSeq 6000Dx produktdokumentation (dokumentnr 200014171) för allmän information om datautmatning och lagring.

TruSight Whole Genome Analysis Application skriver ut data till en Run Folder (körningsmapp) och en Analysis Folder (analysmapp) i extern lagring. Minimilagringskraven kan uppskattas från storleken på datautmatningen till varje mapp för en enda sekvenseringskörning som visas nedan.

Konfiguration	Körningsmapp (GB)	Analysmapp (GB)
S2-flödescell (6 prover)	ca 430	ca 350
S4-flödescell (16 prover)	ca 1 110	ca 890

Ungefärlig analystid

Analysen startar automatiskt efter att en sekvenseringskörning har slutförts och sker sekventiellt på prover inom en körning. Utdatafiler kommer att finnas tillgängliga på den externa lagringsplatsen när analysen är klar för alla prover i en körning och kopieringsöverföringen till den externa lagringsplatsen är klar. När du startar en sekvenseringskörning på både sida A och sida B samtidigt kommer sekvensering att utföras samtidigt. Analys av dessa sekvenseringskörningar kommer att utföras sekventiellt av TruSight Whole Genome Analysis Application efter att sekvenseringen har slutförts. Den körning som slutför sekvensering och överföring först kommer att analyseras först. Den andra sekvenseringskörningen kommer att överföras och placeras i kö för analys efter att den första analysen har slutförts. Se [Visa körning och resultat på sidan 7](#) för information om hur du fastställer status för aktiva eller underkända körningar.

Ungefärlig tid tills analysresultaten är tillgängliga efter att sekvenseringen har slutförts visas nedan för situationen när sida A och sida B laddas samtidigt med samma konfiguration.

Konfiguration	Analyskörning 1 (timmar)	Analyskörning 2 (timmar)
S2-flödescell (6 prover)	ca 12	ca 24
S4-flödescell (16 prover)	ca 24	ca 48

Inställningar

Välj TruSight Whole Genome Analysis Application på skärmen Applications (Applikationer) för att visa aktuell konfiguration och ändra behörigheter.

Konfiguration

Konfigurationsskärmen visar följande programinställningar:

- **Applikationsnamn**
- **Applikationsversion**
- **DRAGEN Version**
- **RTA-version**
- **Datum för utgåva**
- **Organisation**
- **Enhetsidentifierare**
- **Produktionsidentifierare**
- **Library Prep Kits** (Biblioteksprepareringssatser) – Visar biblioteksprepareringssats. Denna inställning kan inte ändras.
- **Index Adapter Kits** (Indexadaptersatser) – Visar de indexadaptersatser som är tillgängliga för användning.

- **Indexavläsningar**
- **Avläsningstyp**
- **Indexlängder**
- **Read Lengths** (Avläsningsslängder) – Avläsningsslängder ställs in som standard när indexuppsättningen väljs. Denna inställning kan inte ändras.

Behörigheter

Den utsedda administratören har behörighetsåtkomst och kan använda kryssrutorna på skärmen Permissions (Behörigheter) för att hantera användaråtkomst för TruSight Whole Genome Analysis Application.

Mer information om behörigheter och användarhantering finns i avsnittet Systemkonfiguration i NovaSeq 6000Dx produktdokumentation (dokument nr 200010105).

Skapande av körning

Skapa nya körningar i IVD-läge, antingen på instrumentet eller genom att öppna Illumina Run Manager (IRM) med en webbläsare på en nätverksansluten dator. För att komma åt instrumentet på distans ska du använda adress- och användarkontoinformationen från din Illumina-representant. Se NovaSeq 6000Dx produktdokumentation (dokument nr 200010105) för mer information.

Create run (Skapa körning) är den rekommenderade metoden för körningsplanering. Import Sample Sheet (Importerera provark) rekommenderas inte. Provarkfilerna som matas ut i körnings- och analysmappar är inte lämpliga för import under körningsplanering.

Skapa körningar

1. På skärmen Runs (Körningar) väljer du **Create Run** (Skapa körning).
2. Välj TruSight Whole Genome Analysis Application och välj sedan **Next** (Nästa).
3. På skärmen Run Settings (Kör inställningar) anger du ett körningsnamn. Körningsnamnet identifierar körningen från sekvensering till analys.
4. [Valfritt] Ange en körningsbeskrivning för att ytterligare identifiera körningen. Biblioteksprepareringsatts är inställd som standard som TruSight Whole Genome och kan inte ändras.
5. Välj önskad TruSight Whole Genome-indexuppsättning i listrutan **Index Adapter Kit** (Indexadaptersats).
Avläsningslängden ställs in som standard och kan inte ändras. (Read 1 (Avläsning 1) och Read 2 (Avläsning 2) använder 151 cykler; Index 1 och 2 använder 10 cykler).
6. Ange ett biblioteksror-ID (rekommenderat format som DX1234567-LIB) och välj sedan **Next** (Nästa).
Om inget biblioteksror-ID specificeras i detta steg måste den planerade körningen väljas innan förbrukningsmaterial för sekvensering laddas. Om fel biblioteksror-ID anges i detta steg måste den planerade körningen korrigeras innan förbrukningsartiklar laddas. Se [Körningsrevision på sidan 5](#) för protokoll för korrekt körning när förbrukningsmaterial är redo att laddas.
7. På skärmarna Sample data (Provdatab) och Sample Settings (Provinställningar) anges provininformation. Provdatab kan matas in manuellt eller genom att importera en provdatafil. Prov-ID måste vara unikt för varje prov och får endast innehålla alfanumeriska tecken, understreck och bindestreck. Inkludera inte mellanslag. Well Position (Brunnsposition) avser brunnen i format A01 till H04 på indexplattan. Indexsekvensinformationen fylls i automatiskt när indexplattans brunnsposition anges. Kön måste anges som Male (Man), Female (Kvinna) eller Unknown (Okänd). Biblioteksplattans ID och biblioteksbrunnens ID (t.ex. format A01) är obligatoriska fält.

- För att mata in provdata manuellt ska du lägga till rader (till totalt 6 för S2 eller 16 för S4 flödescell) och ange nödvändig information i fälten Sample ID (Prov-ID) och Well Position (Brunnsposition). Information kan också kopieras och klistras in från Excel. Välj **Next** (Nästa). På skärmen Sample settings (Provinställningar) anger du Library Plate ID (Biblioteksplattans ID), Library Well ID (Bibliotekets brunns-ID) och Sex (Kön). Välj **Next** (Nästa).
 - För att importera en provdatafil väljer du **Import Samples** (Importera prover) och laddar upp provdatafilen. Informationen fylls i automatiskt i rader. En mall (*.csv) finns tillgänglig för nedladdning på denna skärm. Välj **Next** (Nästa). På skärmen Sample Settings (Provinställningar) fylls information automatiskt i rader från den importerade provdatafilen. Välj **Next** (Nästa).
8. På skärmen Analysis Settings (Analysinställningar) anger du batchnamnet som registrerats under batch- och körningsplanering.
 9. [Valfritt] Välj flödescelltyp, S2 eller S4.
 10. Bekräfta eller avmarkera kryssrutan för att generera ORA-komprimerade FASTQ:er och välj sedan **Next** (Nästa).

OBS! TruSight Whole Genome Analysis Application genererar ORA-komprimerade FASTQ:er som standard. Om du ändrar den här inställningen ökar storleken på den slutliga datautmatningen.

11. Granska den angivna informationen på skärmen Run Review (Körningsgranskning). Om inga ändringar behövs väljer du **Save** (Spara). Om ändringar behövs väljer du **Back** (Tillbaka) efter behov för att återgå till lämplig skärm.



FÖRSIKTIGHET

TruSight Whole Genome har validerats för 6 prover vid användning av NovaSeq 6000Dx S2-flödescellen och 16 prover vid användning av NovaSeq 6000Dx S4-flödescellen. Se till att rätt antal prover anges för den valda flödescellkonfigurationen.

Körningsrevision

Om ändringar krävs efter att körningen har skapats och innan förbrukningsmaterial laddas för sekvensering, revidera körningarna i IVD-läge antingen på instrumentet eller genom att öppna Illumina Run Manager (IRM) med en webbläsare på en nätverksansluten dator.

1. Välj **Runs** (Körningar).
2. Välj körningsnamnet på fliken Planned Runs (Planerade körningar).
3. Välj **Edit** (Redigera).
4. Uppdatera körnings- eller provinformation efter behov. Ange till exempel eller korriger biblioteks-ID för att matcha det som användes när arbetsflödet slutfördes.
5. Välj **Next** (Nästa) för att köra Run Review (Körningsgranskning).
6. Välj **Save** (Spara).

7. Välj **Exit** (Avsluta).

Återgå till sekvensering i IVD-läge för att upprepa laddning av förbrukningsartiklar. Körningen ska nu markeras automatiskt.

Om du uppdaterar biblioteksror-ID när du laddar förbrukningsartiklar återgår du till Run Selection (Körningsval) i kontrollprogrammet och välj **Refresh** (Uppdatera) för den associerade kolumnen, A eller B. Körningen ska nu markeras automatiskt. Om inte väljer du **Back** (Tillbaka) för att upprepa laddning av förbrukningsmaterial.

Repetera analys

Se avsnittet Troubleshooting (Felsökning) i TruSight Whole Genome Package Insert (dokumentnr 200050132) för att avgöra vilken typ av Requeue Analysis (Repetera analys) som är mest lämplig.

Repetera analys utan några ändringar

1. Välj namnet för Completed Run (Fullbordad körning) för att visa Run Details (Körningsdetaljer).
2. Välj **Requeue Analysis** (Repetera analys).
3. Välj **Requeue analysis with no changes** (Repetera analys utan några ändringar).
4. Ange detaljer i fältet Reanalysis Reason (Orsak till repeterad analys).
5. Välj **Requeue Analysis** (Repetera analys).
6. Lämna sidan och navigera till sidan Active Runs (Aktiva körningar) för att bekräfta att repetitionen pågår.

Repetera analys med ändringar

1. Välj namnet för Completed Run (Fullbordad körning) för att visa Run Details (Körningsdetaljer).
2. Välj **Requeue Analysis** (Repetera analys).
3. Välj **Edit run settings** (Redigera körningsinställningar) och **Requeue Analysis** (Repetera analys).
4. Ange detaljer i fältet Reanalysis Reason (Orsak till repeterad analys).
5. Välj **Requeue Analysis** (Repetera analys).
6. Bekräfta eller uppdatera Run Settings (Körningsinställningar) och välj sedan **Next** (Nästa).
7. Korrigera provinformationen efter behov genom att uppdatera fälten manuellt eller välj **Download Template** (Hämta mall) för att skapa en `sampledata.csv`-fil med aktuell information. Korrigera information och ta bort befintliga rader på fliken Sample Data (Provdatab) innan du använder Import Samples (Importerar prover) för att fylla i korrigerade provdata.
8. Granska informationen på skärmen Run Review (Körningsgranskning) och välj **Save** (Spara) för att starta repeterad analys.
9. Välj **Exit** (Avsluta) och navigera till sidan Active Runs (Aktiva körningar) för att bekräfta att repetitionen pågår.

Den ursprungliga körningsdatamappen måste finnas på den externa lagringsplatsen som anges i Run Details (Körningsdetaljer) för att den repeterade analysen ska kunna slutföras. Om repeterad analys misslyckas ska du se till att körningen inte har flyttats eller tagits bort.

Visa körning och resultat

1. På huvudskärmen för Illumina Run Manager i IVD-läge väljer du **Runs** (Körningar).
2. På fliken Completed Runs (Slutförda körningar) väljer du körningsnamn.
Den här fliken visar även körningar som har slutförts på grund av misslyckad sekvensering, dataöverföring eller analys. Aktiva körningar och deras status visas på fliken Active Runs (Aktiva körningar). Se NovaSeq 6000Dx produktdokumentation (dokument nr 200010105) för mer information.
3. Välj Körningsnamn på fliken Completed Runs (Slutförda körningar) för att visa körningsdetaljer och resultat för sökvägen till analysutdatamappen.
För underkända körningar ska du granska status för varje steg och sedan se avsnittet Troubleshooting (Felsökning) i TruSight Whole Genome Package Insert (dokumentnr 200050132).
4. Navigera till analysmappen på din lokala enhet och öppna Consolidated Report (Konsoliderad rapport) för att granska resultatet GODKÄND/UNDERKÄND för varje QC-steg enligt följande:
 - För kvalitetskontroll av sekvenseringskörning ska du se Sammanfattning av QC-resultat för sekvensering
 - För FASTQ-QC för varje prov i körningen ska du se Sammanfattning av FASTQ QC-resultat
 - För biblioteks-QC för varje prov i körningen ska du se Sammanfattning av QC-resultat för provbibliotekOm ett UNDERKÄND-resultat observeras ska du notera QC-steget och se avsnittet Troubleshooting (Felsökning) i TruSight Whole Genome Package Insert (dokumentnr 200050132).

Sammanfattning av utdatafil

TruSight Whole Genome Analysis Application sparar följande huvudutdatafiler. Se filinformationsavsnitten nedan för placering av huvudutdatafiler.

Körningar och prover som inte uppfyller giltighetskriterierna producerar inte CRAM-, ROH bed- eller *genome.vcf-filer.

Utdatafil	Beskrivning
Konsoliderad rapport (*.csv)	Innehåller kvalitetsmått som används för körningsgiltighet (inklusive totalt utbyte och Q30), provgiltighetsmått (inklusive FASTQ-kapacitet), QC-mått för bibliotek och mått för endast för information (FIO) för alla prover i körningen.
Exempelrapport (*.csv)	Innehåller resultat från QC för sekvensering, FASTQ och QC för provbibliotek. Rapporten innehåller också ploiditetöverensstämmelse och FIO-mått för det enskilda provet samt tillhörande sekvenseringskörning.
Liten variant och mSNV VCF (*.annotated.hard-filtrerad-gvcf.gz)	Innehåller variantbestämningsinformation för små varianter (SNV:er, indeler) och mitokondriella SNV:er.
CNV VCF (*.cnv.vcf.gz)	Innehåller variantbestämningsinformation för kopietalsvinster och -förluster.
Upprepar VCF (*.repeats.vcf.gz)	Innehåller variantbestämningsinformation om STR-expansioner och SMN1.
ROH BED (*.roh.bed)	Innehåller information för regioner med homozygositet.
FASTQ (*.fastq.gz eller *.fastq.ora)	Intermediära filer som innehåller basbestämningar med kvalitetspoäng. FASTQ-filer är inpassningsstegets primära indata. Om ORA-komprimering väljs återspeglar filnamnet detta.
Inpassnings-CRAM (*.cram)	Innehåller justerade läsningar för ett givet prov.

Information om QC-rapport

Katalogen Consolidated Report (konsoliderad rapport) <<RunID>>_Consolidated_Report.csv finns i katalogen TruSightWholeGenomeAnalysis_x.x.x_run-complete och innehåller information om kvalitetsmått som används för att godkänna eller underkänna prover i olika analysstadiet. Enskilda provrapporter <<Sample_ID>>_Sample_Report.csv finns i mapparna <Sample_ID> i katalogen TruSightWholeGenomeAnalysis_x.x.x_run-complete.

Rapportrubrikerna innehåller följande information om körningen: appversionen, batchnamnet, bibliotekspoolrörs-ID, sekvenseringskörningsnamnet, sekvenseringskörnings-ID och flödescelltyp. Följande tabeller beskriver informationen som ingår i den Consolidated Report (konsoliderad rapport). Den individuella provrapporten innehåller samma information förutom demultiplexmått.

Tabell 1 QC-mått för sekvensering

Mätvärden	Specifikationer	Beskrivning
Icke-indexerat totalt utbyte (GB)	N/A (Ej tillämpligt)	Ingen specifikation eftersom körning med lägre utbyte kan leda till att provbiblioteken godkänns. Räkna med $\geq 3\ 000$ Gbp för S4 och $\geq 1\ 000$ GB för S2 flödescell.
Totalt % \geq Q30	≥ 85	Mätning av baskvalitet på körnivå. Minsta specifikation ställs in eftersom alltför låga %Q30-körningar inte klarar Q30-baser i QC av provbibliotek.
Sammanfattning av QC-resultat för sekvensering	GODKÄND eller UNDERKÄND	För underkänd QC för sekvensering ska du se avsnittet Troubleshooting (Felsökning) i TruSight Whole Genome Package Insert (dokumentnr 200050132).

Tabell 2 Demultiplex-mätvärden

Mätvärden	Specifikationer	Beskrivning
Procent läsningar identifierade	N/A (Ej tillämpligt)	Total andel godkända filteravläsningar i körningen som tilldelades till prover under demultiplexering.
Procent CV	N/A (Ej tillämpligt)	Ger ett mått på jämnheten hos avläsningar som demultiplexerats till varje indexpar på körningen. Förvänta dig $< 25\ %$ för körningar utan underkända FASTQ QC.

Tabell 3 FASTQ QC-mått

Mätvärden	Specifikationer	Beskrivning
Utbyte per prov (bps)	$\geq 90\ 000\ 000\ 000$	Minimum är inställt på att motsvara ca 26 x genomsnittlig autosomal täckning för att triagera provbibliotek som underkänns i QC för att minska analystiden.
Sammanfattning av FASTQ QC-resultat	GODKÄND eller UNDERKÄND	För underkänd QC för FASTQ ska du se avsnittet Troubleshooting (Felsökning) i TruSight Whole Genome Package Insert (dokumentnr 200050132).

Tabell 4 QC-mått för provbibliotek

Mätvärden	Specifikationer	Beskrivning
Genomsnittlig autosomal täckning	≥ 35	Genomsnittlig täckning över autosomerna. Minsta specifikation ställs in för att säkerställa analytisk prestanda.
Procentandel autosomer med mer än 20 x täckning	$\geq 93,94$	Mätning av täckningens enhetlighet som upptäcker problem som inte nödvändigtvis är relaterade till GC-bias. Minsta specifikation ställs in för att säkerställa analytisk prestanda.
Normaliserad täckning vid 60 % till 79 % GC-grupper	$0,82 \leq x \leq 1,13$	Mätning av täckningsenlighet som detekterar GC-bias, specifikt en förlust av täckning i områden av genomet med högre % GC och lägre % AT-bassammansättning. Minsta och maximala specifikationer ställs in för att säkerställa analytisk prestanda.
Normaliserad täckning vid 20 % till 39 % GC-grupper	$0,97 \leq x \leq 1,06$	Mätning av täckningsenlighet som detekterar GC-bias, specifikt en förlust av täckning i områden av genomet med lägre % GC och högre % AT-bassammansättning. Minsta och maximala specifikationer ställs in för att säkerställa analytisk prestanda.
Genomsnittlig mitokondriell täckning	≥ 500	Täckning av mitokondriell kromosom. Minsta specifikation ställs in för att säkerställa detektionsgränsen för mitokondriell SNV.
Procent Q30-baser	≥ 85	Mätning av baskvalitet. Minsta specifikation ställs in för att säkerställa analytisk prestanda.
Uppskattad provkontaminering	$\leq 0,005$	Detekterar kontaminerande avläsningar från andra prover. Maximal specifikation ställs in för att säkerställa detektionsgränsen för mitokondriell SNV (varianttypen med den högsta känsligheten för kontaminering).
Sammanfattning av QC-resultat för provbibliotek	GODKÄND eller UNDERKÄND	För underkänt QC för provbibliotek ska du se avsnittet Troubleshooting (Felsökning) i TruSight Whole Genome Package Insert (dokumentnr 200050132).

Tabell 5 QC-mått för ploiditet

Mätvärden	Specifikationer	Beskrivning
Angiven könskromosomploiditet	N/A (Ej tillämpligt)	Kön som anges av operatören under körning (kvinna, man, okänd).
Ploiditetsuppskattning	N/A (Ej tillämpligt)	Könsploiditet uppskattas till DRAGEN.
Sammanfattning av ploiditetsresultat	CONCORDANT (ÖVERENSSTÄMMANDE), DISCONCORDANT (ICKE-ÖVERENSSTÄMMANDE) eller ND (EJ BESTÄMD)	CONCORDANT (ÖVERENSSTÄMMANDE) anger överensstämmande mellan angiven och uppskattad könsploiditet. ND (EJ BESTÄMD) anger att kön som angetts som okänt eller annan uppskattning än XX eller XY. För resultat som är DISCORDANT (ICKE-ÖVERENSSTÄMMANDE) angavs antingen fel kön när körningen skapades eller så kan provbyte ha inträffat. Se avsnittet Troubleshooting (Felsökning) i TruSight Whole Genome Package Insert (dokumentnr 200050132).

Tabell 6 Mått som är endast för information

Mätvärden	Beskrivning
Medianlängd för insertion	Målet är 450 bp men förvänta en variation beroende på sekvenseringskörning och operatör. Ett intervall på cirka 360 till 550 bp är acceptabelt. Konsekvent drift utanför detta intervall kan leda till en högre förekomst av felaktiga prov.
Procent mappade avläsningar	Procentandel av avläsningarna som mappar till referensgenomet. Kan minska som svar på icke-mänsklig gDNA-kontamination, dålig provkvalitet eller för liten insertionslängd, vilket leder till mappningsproblem.
Procentavläsningar med kompletterande inpassningar	Procent av avläsningar med mappning som delas över olika platser i referensgenomet.
Procent dubletter av markerade avläsningar	Förvänta dig < 20 %. Kan höjas om biblioteksprepareringsutbytet är lågt eller poolar mindre än krävd volym, eller som svar på sekvenseringsrelaterade problem.

Mätvärden	Beskrivning
Procent mjuka klippta baser Read 1 (Avläsning 1)	Användbar vid diagnos av grundorsak för fel på genomsnittlig autosomal täckning.
Procent mjuka klippta baser Read 2 (Avläsning 2)	Användbar vid diagnos av grundorsak för fel på genomsnittlig autosomal täckning.
Procent baser, trimmade Read 1 (Avläsning 1)	Användbar vid diagnos av grundorsak för fel på genomsnittlig autosomal täckning.
Procent baser, trimmade Read 2 (Avläsning 2)	Användbar vid diagnos av grundorsak för fel på genomsnittlig autosomal täckning.

Information om variantbestämningsfil

VCF-filer

Filer med variantbestämningsformat (*.vcf) innehåller information om varianter som finns på specifika positioner i ett referensgenom och återfinns i katalogen <Sample_ID>/Analysis.

VCF-filhuvudet inkluderar VCF-filformatversionen och variantbestämningsversionen och listar de anteckningar som används i resten av filen. Den sista raden i huvudet innehåller kolumnhuvudet för datalinjerna. Var och en av VCF-filens datarader innehåller information om en enda referensposition.

Alla VCF-filer innehåller en rubrik med beskrivningar av utdatakolumner och variantbestämningsdata i kolumner märkta som CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO, FORMAT, SAMPLE. Definitioner av kolumnvärden kan variera mellan variantbestämningar.

Liten variant och mSNV VCF

Utdata sparas under filen <Sample_ID>.annotated.hard-filtered.gvcf.gz i katalogen <Sample_ID>/Analysis.

En genomisk VCF-fil (gVCF) innehåller information om varianter och positioner som fastställts vara homozygota för referensgenomet. För homozygota regioner innehåller gVCF-filen statistik som anger hur väl läsningar stöder frånvaron av varianter eller alternativa alleler. gVCF-filen innehåller en konstgjord <NON_REF>-allel. Avläsningar som inte stöder referensen eller några varianter tilldelas <NON_REF>-allelen. DRAGEN använder dessa avläsningar för att avgöra om positionen kan bestämmas som en homozygot referens, i motsats till att förbli obestämd. Den resulterande poängen representerar den Phred-skalade nivån för förtroende för en homozygot referensbestämning. I könscealläge är poängen FORMAT/GQ.

DRAGEN tillhandahåller variantfiltrering efter VCF baserat på kommentarer som finns i VCF-posterna. Hårdfiltrering av variant beskrivs nedan. På grund av arten av DRAGEN:s algoritmer, som införlivar hypotesen om korrelerade fel inifrån kärnan av variantbestämning, har arbetsflödet förbättrats med avseende på förmågan att skilja de sanna varianterna från brus och därför minskar beroendet av filtrering efter VCF avsevärt.

TruSight Whole Genome Analysis Application ger en annotering om konfidenspoäng och konfidensnivå för små varianter som kan användas för att ytterligare förbättra prestandan. Förtroendenivåannotering är inte ett kvalitetsfilter och återspeglas därför inte direkt i variantbestämningarnas kvalitetsstatus. Därför är det möjligt att se passerande variantbestämningar som ändå är annoterade som lågt förtroende.

Tabell 7 VCF-filrubriker

Rubrik	Beskrivning
CHROM	Referensgenomets kromosom. Kromosomerna visas i samma ordning som referens-FASTA-filen.
POS	Variantens enkelbasposition i referenskromosomen. För enkelnukleotidvarianter (SNV) är denna position referensbasen med varianten. För indels är denna position referensbasen omedelbart före varianten.
ID	Alltid .
REF	Referensgenotypen. Till exempel representeras en deletion av ett enda T som referens TT och alternativ T. En enkel nukleotidvariant A till T representeras som referens A och alternativ T.
ALT	Allelerna som skiljer sig från referensen läses. Till exempel representeras en insertion av ett enda T som referens A och alternativ AT. En enkelnukleotidvariant A till T representeras som referens A och alternativ T.
QUAL	En Phred-skalad kvalitetspoäng tilldelad av variantbestämaren. Högre poäng indikerar högre tilltro till varianten och lägre sannolikhet för fel. För kvalitetspoäng Q är den uppskattade sannolikheten för ett fel $10^{-(Q/10)}$. Till exempel har uppsättningen Q30-bestämningar en felfrekvens på 0,1 %. Många variantbestämare tilldelar kvalitetspoäng baserat på deras statistiska modeller, som är höga i förhållande till den observerade felfrekvensen.

Tabell 8 VCF-filanteckningar

Rubrik	Beskrivning
FILTER	<ul style="list-style-type: none"> • GODKÄND: Alla filter är godkända. • DRAGENSnpHardQUAL – Ställ in om true:QUAL < 10,41 • DRAGENIndelHardQUAL – Ställ in om true:QUAL < 7,83 • LowDepth – Ställ in om true:DP <= 1 • LowGQ – Ställ in om true:GQ = 0 • PloidyConflict – Genotypbestämning från variantbestämmare överensstämmer inte med kromosomploidi • base_quality – Platsen filtreras eftersom medianbaskvaliteten för alt-läsningar på detta lokus inte når tröskeln • filtered_reads – Platsen har filtrerats eftersom en för stor del av läsningarna har filtrerats bort • fragment_length – Platsen filtreras eftersom den absoluta skillnaden mellan medianfragmentlängden för alt-läsningar och medianfragmentlängden för ref-läsningar på detta ställe överskrider tröskeln • low_af – Allelfrekvensen uppfyller inte tröskelvärdet • low_depth – Platsen filtreras eftersom läsdjupet är för lågt • low_frac_info_reads – Platsen filtreras eftersom andelen informativa läsningar ligger under tröskeln • low_normal_depth – Platsen filtreras eftersom det normala provets läsdjup är för lågt • long_indel – Platsen filtreras eftersom indellängden är för lång • mapping_quality – Platsen filtreras eftersom medianmappningskvaliteten för alt-läsningar på detta lokus inte når tröskeln • multiallelic – Platsen filtreras eftersom mer än två alt-alleler passerar tumör-LOD • non_homref_normal – Platsen filtreras eftersom den normala provgenotypen inte är homozygot referens • no_reliable_supporting_read – Platsen filtreras eftersom det inte finns någon pålitlig stödjande somatisk läsning • panel_of_normals – Ses i minst ett prov i panelen av normaler vcf • read_position – Platsen filtreras eftersom medianen för avstånden mellan start/slut av läsning och detta lokus ligger under tröskeln • RMxNRepeatRegion – Platsen filtreras eftersom hela eller delar av variantallelen är en upprepning av referensen • strand_artifact – Platsen filtreras på grund av allvarlig strängbias • str_contraction – Platsen filtrerad på grund av misstänkt PCR-fel där alt-allelen är en upprepad enhet mindre än referensen • too_few_supporting_reads – Platsen filtreras eftersom det finns för få stödjande avläsningar i tumörprovet • weak_evidence – Somatisk variantpoäng når inte tröskeln

Rubrik	Beskrivning
INFO	<ul style="list-style-type: none"> • DB – dbSNP-medlemskap. • FS – Phred-skalat p-värde med Fishers exakta test för att upptäcka strängförspänning. • QD – Varianttillförlitlighet/-kvalitet efter djup. • R2_5P_bias – Poäng baserat på parningspartiskhet och avstånd från 5 primeslut. • SOR – Symmetrisk oddskvot på 2x2 beredskapstabell för att upptäcka strängförspänning. • DP – Ungefärligt läsdjup (informativt och icke-informativt); vissa läsningar kan ha filterats baserat på mapq etc. • END – Stopp position för intervallet. • FractionInformativeReads – Bråkdelen av informativa läsningar av det totala antalet läsningar. • MQ – RMS Kartläggningsskvalitet. • MQRankSum – Z-poäng från Wilcoxon rangsummetest av Alt vs Ref läs mappningskvaliteter. • ReadPosRankSum – Z-poäng från Wilcoxon rangsummetest av Alt vs. Ref läs positionsbias. • SOMATIC – Minst en variant på denna position är somatisk. • ILENS – Indellängder för varje ALT-variant. • SCORE – Förtroendepoäng för varje varianttyp som finns på platsen som (variant type):(confidence score). • TIER – Förtroendenivå för varje varianttyp som finns på platsen som (variant type):(confidence tier).

Rubrik	Beskrivning
FORMAT	<p>Kolumnen FORMAT visar fält separerade med kolon, till exempel GT:QC.</p> <ul style="list-style-type: none"> • AD – Alleldjup (räknas endast informativa läsningar av de totala läsningarna) för ref- och alt-allelerna i den ordning som anges. • AF – Allelfraktioner för alt-alleler i den ordning som anges. • DP – Ungefärligt läsdjup (avläsningar med MQ=255 eller med dåliga kompisar filtreras). • F1R2 – Antal avläsningar i F1R2-parorientering som stöder varje allel. • F2R1 – Antal avläsningar i F2R1-parorientering som stöder varje allel. • GP – Phred-skalade posteriora sannolikheter för genotyper enligt definitionen i VCF-specifikationen. • GQ – Genotypkvalitet. • GT – Genotyp. • ICNT – Antal informativa INDEL-avläsningar baserade på referensförtroendemodellen. • MB – Per-prov komponentstatistik för att upptäcka parningspartiskhet. • MIN_DP – Minsta DP observerad inom GVCF-blocket. • PL – Normaliserade, Phred-skalade sannolikheter för genotyper enligt definitionen i VCF-specifikationen. • PRI – Phred-skalade tidigare sannolikheter för genotyper. • PS – Fysisk fas-ID-information, där varje unikt ID inom ett givet prov (men inte över prover) kopplar samman poster inom en fasningsgrupp. • SB – Per-prov komponentstatistik som omfattar Fishers Exact Test för att upptäcka strängbias. • SPL – Normaliserade, Phred-skalade sannolikheter för SNP:er baserat på referensförtroendemodellen. • SQ – Somatisk kvalitet.
PROV	Exempelkolumnen ger de värden som anges i FORMAT-kolumnen.

Kopietalsvariation (VCF)

Målräkningssteget är det första bearbetningssteget för DRAGEN CNV-arbetsflöde och som producerar `<Sample_ID>.target.counts.gz`, sedan utförs GC Bias Correction vilket genererar en `*.target.counts.gc-corrected.gz`-fil. Normaliseringssteget producerar `*.tn.tsv.gz`-filen. Världprogramvaran DRAGEN genererar många mellanliggande filer. `*.seg.called.merged` är den slutliga bestämningsfilen som innehåller amplifierings- och deletionshändelserna.

Förutom segmentfilen avger DRAGEN bestämningarna i standardiserat VCF-format. Utdata sparas i `<Sample_ID>.cnv.vcf.gz` i katalogen `<Sample_ID>/Analysis`.

Definitioner av kolumner som är specifika för CNV-bestämmare:

POS-kolumnen är variantens startposition. Enligt VCF-specifikationen, om någon av ALT-allelerna är en symbolisk allel, såsom , krävs vadderingsbasen och POS betecknar koordinaten för basen som föregår polymorfismen. Alla koordinater i VCF är 1-baserade.

ID-kolumnen används för att representera händelsen. ID-fältet kodar händelsetypen och koordinaterna för händelsen.

REF-kolumnen innehåller ett N för alla CNV-händelser.

ALT-kolumnen anger typen av CNV-händelse. Eftersom VCF endast innehåller CNV-händelser används endast posten DEL eller DUP.

QUAL-kolumnen innehåller en uppskattad kvalitetspoäng för CNV-händelsen, som används vid hård filtrering.

FILTER-kolumnen innehåller PASS om CNV-händelsen passerar alla filter, annars innehåller kolumnen namnet på det underkända filtret.

INFO-kolumnen innehåller information som representerar händelsen. Posten REFLLEN anger händelsens längd. Posten SVTYPE är alltid CNV. Posten END anger händelsens slutposition.

FORMAT-fälten beskrivs i rubriken.

- GT – Genotyp
- SM – Linjärt kopieförhållande för segmentets medelvärde
- CN – Uppskattat kopietal
- BC – Antal grupper i regionen
- PE – Antal felaktiga avläsningar med parad ände vid start och stopp av brytpunkter

Upprepar VCF

ExpansionHunter utför en sekvensgrafbaserad upprepad inpassning av avläsningar som har sitt ursprung inuti och runt varje måluppreppning. ExpansionHunter genotyper sedan längden på uppreppningen i varje allel baserat på dessa grafpassningar.

Mer information och analys finns i följande ExpansionHunter-dokument:

- Dolzhenko et al., *Detection of long repeat expansions from PCR-free whole-genome sequence data* 2017
- Dolzhenko et al., *ExpansionHunter: A sequence-graph based tool to analyze variation in short tandem repeat regions* 2019

STR-variantkatalogen i TruSight Whole Genome Analysis Application innehåller specifikationer för sjukdomsorsakande uppreppningar i generna AFF2, AR, ATN1, ATXN1, ATXN10, ATXN2, ATXN3, ATXN7, ATXN8OS, C9ORF72, CACNA1A, CBL, CNBP, CSTB, DIP2B, DMPK, FMR1, FXN, GLS, HTT, JPH3, NIPA1, NOP56, NOTCH2NL, PABPN1, PHOX2B, PPP2R2B och TBP.

Resultaten av upprepad genotypbestämning matas ut som en separat VCF-fil och som ger längden på varje allel vid varje bestämningsbar upprepning som definieras i katalogfilen för upprepningsspecifikation. Filnamnet är <Sample_ID>.repeats.gz och finns i katalogen <Sample_ID>/Analysis.

Vissa kolumner är specifika för att upprepa expansionsbestämning:

Tabell 9 Huvudsakliga VCF-fält

Fält	Beskrivning
CHROM	Kromosomidentifikatorer
POS	Position för den första basen före det upprepade området i referensen
ID	Alltid .
REF	Referensbasen vid position POS
ALT	Lista över upprepade alleler i formatet <STRn>. N är antalet upprepade enheter.
QUAL	Alltid .
FILTER	LowDepth-filter används när det totala locusdjupet är under 10 x eller antalet avläsningar som spänner över en eller båda brytningarna är under 5.

Tabell 10 Ytterligare INFO-fält

Fält	Beskrivning
END	Position för den sista basen för det upprepade området i referensen
REF	Referenskopietal
REPID	Variant-ID från variantkatalogen
RL	Referenslängd i bp
RU	Upprepad enhet i referensorienteringen
VARID	Variant-ID från variantkatalogen

Tabell 11 Fält för GENOTYPE (per prov)

Fält	Beskrivning
AD	Alleldjup för ref- och alt-allelerna i den ordning som anges
ADFL	Antal flankerande avläsningar som överensstämmer med allelen
ADIR	Antal in-upprepade avläsningar som överensstämmer med allelen
ADSP	Antal överspännande avläsningar som överensstämmer med allelen
DST	Resultat (+ detekterade, - oupptäckta, ? obestämda) för testet representerat av varianten

Fält	Beskrivning
GT	Genotyp
LC	Locustäckning
REPCI	Konfidensintervall för REPCN
REPCN	Antal upprepade enheter som överspänns av allelen
RPL	Förhållande mellan logg-sannolikhet för närvaro av referensallelen
SO	Typ av avläsningar som stöder allelen. Värdena kan vara <code>SPANNING</code> (ÖVERSPÄNNANDE), <code>FLANKING</code> (FLANKERANDE) eller <code>INREPEAT</code> (IN-UPPREPADE). Dessa värden indikerar om avläsningen överspänner, flankerar eller är helt inneslutna i upprepningen.

Filen `<Sample_ID>.repeats.vcf.gz`-filen innehåller SMN-utdata tillsammans med eventuella målupprepningar. SMN-utdata representeras som en enda SNV-bestämning vid splicepåverkande position i SMN1 (NM_000344.3:c.840C/T) med status för spinal muskelatrofi (SMA) i följande anpassade fält.

Tabell 12 SMA-resultat i utdatafil repeats.vcf

Fält	Beskrivning
VARID	SMN markerar SMN-bestämningen.
GT	Genotypbestämning vid denna position med en normal (diploid) genotypmodell.
DST	SMA-statusbestämning: + indikerar detekterad - indikerar odetekterad ? indikerar obestämmd
AD	Totalt antal avläsningar som stöder C- och T-allelen.
RPL	Log ₁₀ -sannolikhetskvot mellan opåverkade och påverkade modeller. Positiva poäng indikerar att den opåverkade modellen är mer sannolik.

ROH BED

Regioner av homozygositet (ROH) detekteras som en del av bestämningen av små varianter. Bestämningen upptäcker och ger utdata för homozygositetskörningar från hela genombestämningar på autosomala mänskliga kromosomer. Könskromosomer ignoreras såvida inte provets könskaryotyp är XX, vilket bestäms av Ploidy Estimator. ROH-utdata gör det möjligt för verktyg nedströms att screena för och förutsäga konsangvinitet mellan föräldrarna till proband-personen.

En region definieras som på varandra följande variantbestämningar på kromosomen utan någon stor lucka mellan dessa varianter. Med andra ord bryts regioner av kromosomer eller av stora luckor utan SNV-bestämningar. Luckstorleken är inställd på 3 Mbaser.

ROH-bestämningen producerar en ROH-utdatafil med namnet `<Sample_ID>.roh.bed` i katalogen `<Sample_ID>/Analysis`. Varje rad representerar en region av homozygositet. Bed-filen innehåller följande kolumner:

```
Chromosome Start End Score #Homozygous #Heterozygous
```

Var

- Poäng är en funktion av antalet homozygota och heterozygota varianter, där varje homozygot variant ökar poängen med 0,025, och varje heterozygot variant minskar poängen med 0,975.
- Start- och slutpositioner är ett 0-baserat halvöppet intervall.
- #Homozygot är antalet homozygota varianter i regionen.
- #Heterozygot är antalet heterozygota varianter i regionen.

Bestämningen producerar också en mätvärdesfil med namnet `<Sample_ID>.roh_metrics.csv` som listar antalet stora ROH och procentandelen SNP i stora ROH (> 3 MB).

Uppskattningsmått för ploiditet

Ploidy Estimator körs som standard. Ploidy Estimator använder avläsningar från mapparen/inpassaren för att beräkna sekvenseringsdjupet för täckning för varje autosom och allosom i det mänskliga genomet. Könskaryotypen för provet uppskattas sedan med hjälp av kvoterna mellan medianvärdet för könskromosomtäckning och medianvärdet för autosomal täckning. XX eller XY, och CONCORDANT (ÖVERENSSTÄMMANDE), DISCORDANT (ICKE-ÖVERENSSTÄMMANDE) eller ND (EJ FASTSTÄLLD) jämfört med provdata som tillhandahålls rapporteras i den konsoliderade rapporten. De detaljerade resultaten, inklusive varje normaliserad mediantäckning per contig, rapporteras i filen `<Sample_ID>.ploidy_estimation_metrics.csv`.

FASTQ-filer

FASTQ (*.fastq.gz, *.fastq.ora) är ett textbaserat filformat som innehåller basbestämningar och kvalitetsvärden per läsning. Varje fil innehåller följande information:

- Providentifierare
- Sekvensen
- Ett plustecken (+)
- Phred-kvalitetspoäng i ett ASCII + 33-kodat format

Programvaran genererar en FASTQ-fil för varje prov, avläsning och spår. För varje prov i en körning av parad ände genererar programvaran till exempel två FASTQ-filer: en för Read 1 (Avläsning 1) och en för Read 2 (Avläsning 2). Förutom dessa exempel på FASTQ-filer genererar programvaran två FASTQ-filer per spår som innehåller alla okända prover. FASTQ-filer för Index Read 1 (Indexavläsning 1) och Index Read 2 (Indexavläsning 2) genereras inte eftersom sekvensen ingår i rubriken för varje FASTQ-post. Filnamnsformatet skapas från fält som specificeras i provarket och använder filnamnsformatet

```
<Sample_ID>_S#_L00#_R#_001.fastq.gz
```


FASTQ-filer sparas i katalogen `<Sample_ID>/Conversion`. I FASTQ-katalogen i analysmappen kan man hitta loggkatalogen med BCL-till-FASTQ-konverteringsloggar och rapportkatalogen som innehåller olika avläsningsmätvärdesfiler och `SampleSheet.csv` som används för FASTQ-konvertering. FASTQ-filer från obestämda avläsningar finns i katalogen `Undetermined/Conversion` (Obestämd/Konvertering) i analysmappen.

Providentifieraren formateras enligt följande:

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber
Exempel:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=#
```

CRAM-filer

Komprimerad referensorienterad inpassningskarta (CRAM-filer) (*.cram) lagras i katalogen `<Sample_ID>/Analysis` och innehåller rubriker och inpassningsposter i förhållande till den genomiska referensfilen som används under inpassning. Sökvägen till referensfilen visas i filen `<Sample_ID>/Analysis/<Sample_ID>-replay.json` som en `--ht-reference` parameter, som är standardinställd på `hg38.fa`.

CRAM-filer innehåller en rubriksektion och en inpassningssektion:

- **Header** (Rubrik) – Innehåller information om hela filen, som provnamn, provlängd och justeringsmetod. Justeringar i inriktningssektionen är associerade med specifik information i rubriksektionen.
- **Alignments** (Justeringar) – Innehåller läsnamn, lässekvens, läskvalitet, anpassningsinformation och anpassade taggar. Läsnamnet inkluderar kromosomen, startkoordinaten, justeringens kvalitet och matchningsbeskrivningssträngen.

Justeringssektionen innehåller följande information för varje läs- eller läspar:

- AS: Inriktningskvalitet med parad ände.
- RG: Läsgrupp, som anger antalet läsningar för ett specifikt prov.
- BC: Streckkodstagg, som indikerar det demultiplexade prov-ID som är kopplat till läsningen.
- SM: Ensidig inriktningskvalitet.
- XC: Matcha deskriptorsträng.
- XN: Tagg för amplikonnamn, som registrerar det amplikon-ID som är kopplat till läsningen.

För att visa inpassningsposter kan `sam tools` (samverktyg) användas som `samtools view --reference <path_to_reference_folder>/hg38.fa <Sample_ID>.cram`.

En indexfil och en kontrollsummafil genereras också.

Teknisk hjälp

För teknisk hjälp kontaktar du Illumina tekniska support.

Webbplats: www.illumina.com

E-post: techsupport@illumina.com

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

Produktdokumentation — Kan hämtas på support.illumina.com.



Illumina, Inc.
5200 Illumina Way
San Diego, California 92122 USA
+1 800 809 ILMN (4566)
+1 858 202 4566 (utanför Nordamerika)
techsupport@illumina.com
www.illumina.com



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

Australisk sponsor

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

FÖR IN VITRO-DIAGNOSTISKT BRUK.

© 2024 Illumina, Inc. Med ensamrätt.

illumina[®]